



The Structure and Development of Logical Representations in Thought and Language

Citation

Feiman, Roman. 2015. The Structure and Development of Logical Representations in Thought and Language. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:23845487>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

The structure and development of logical representations in thought and language

A dissertation presented

by

Roman Feiman

to

The Department of Psychology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University

Cambridge, Massachusetts

September, 2015

© Roman Feiman
All rights reserved.

Dissertation Advisors: Professor Susan Carey
Professor Jesse Snedeker

Roman Feiman

The structure and development of logical representations in thought and language

Abstract

The expressive power of human thought and language comes from the ability to systematically combine a finite vocabulary of concepts into a boundless number of meaningful thoughts. What properties of conceptual representations enable their combination? Three papers investigate different aspects of the combinatorial system in the context of a single general approach – taking logical concepts as a special case of concepts whose content is completely specified by their combinatorial properties. The first paper looks at infants’ ability to represent two types of goals: approach and avoid, where each goal-type could be represented as the negation of the other. Consistent with past literature, we find evidence of children representing approach at 7 month, but failing to represent avoid at both 7 and 14 months. This suggests that these children cannot combine their representation of approach with a negation operator, possibly because they do not yet have this operator. In the second paper, we continue to look at the emergence of logical negation through the relationship between the emergence of the concept and the words that label it. We find that, although 15-month-olds say the word “no”, they do not understand its logical meaning until 24 months. This is the same age at which they begin to produce the word “not”, comprehend its logical meaning, and use both “no” and “not” to deny the truth of others’ statements. This pattern of results suggest a common limiting factor on the mapping of any word to the concept of logical negation. This factor could be the emergence of

the concept, or a linguistic limitation common to both “no” and “not”. The third paper looks at the properties of the combinatorial system in adults, taking linguistic quantifier scope ambiguity phenomena as a case study. Using a priming paradigm, we find evidence for independent combinatorial operations for the universal quantifiers EACH, EVERY and ALL, but common operations for the numbers THREE, FOUR and FIVE. We also find that the semantic operations that compose quantifier meanings abstract away from the verb and noun content of sentences. This suggests a division of labor in adult combinatorial thought, with conceptual content represented separately from the combinatorial properties of concepts.

Table of Contents

Chapter 1: Introduction.....	1
1.1. Combinatorial thought.....	2
1.2. The relationship between format and content.....	3
1.3. The way forward.....	13
1.4. The three papers.....	17
1.5. The end of the beginning.....	23
Chapter 2: Infants' representations of others' goals: Representing approach over avoidance.....	25
2.1. Introduction.....	26
2.2. Experiment 1.....	30
2.3. Experiment 2.....	37
2.4. Experiment 3.....	41
2.5. General Discussion.....	47
Chapter 3: What do you mean, “No”? Young children’s production of some negation words precedes their understanding of logical negation.....	54
3.1. Introduction.....	55
3.2. Experiment 1.....	65
3.3. Experiment 2.....	76
3.4. Experiment 3.....	80
3.5. Experiment 4.....	93
3.6. General Discussion.....	99
Chapter 4: The logic in language: How all quantifiers are alike, but each quantifier is different.....	111
4.1. Introduction.....	112
4.2. Experiment 1.....	121
4.3. Experiment 2.....	126
4.4. Experiment 3.....	142
4.5. Experiment 4.....	146
4.6. General Discussion.....	148
4.7. Conclusion.....	159
Chapter 5: Conclusion.....	161
5.1. The development of negation.....	162
5.2. The development of quantification.....	166
5.3. Content and combination.....	169
References.....	170

Acknowledgements

I owe the greatest thanks to my advisors, Susan Carey and Jesse Snedeker. It is hugely rewarding to work with either of them; that I have gotten to work with both seems unfairly lucky on my part. Neither this work, nor any work I do in the future, would be possible without their guidance, support, wisdom and infinite patience. Once I've hatched, and am trying to do some passable research without the benefit of our weekly meetings, I will constantly be asking, "What would Jesse and Susan think of this?" And both I and the work would always benefit from knowing the answer. Most importantly, I cannot imagine better models and mentors in the most important academic skills I know: decency and kindness. Thank you.

I also owe many thanks to the other two members of my committee, Liz Spelke and Steve Pinker. Reading Liz's papers as an undergraduate turned me onto developmental psychology. What I didn't know then is that every time Liz goes to a talk or comments on a paper, she makes them better. I have been very lucky to be able to discuss this work with her. For all the very nice things that have been said about Steve, he may still not get enough credit for being among the most responsive, generous and supportive people in our field. His contributions to this work extend from helpful conversations about the elements of both style and substance, all the way back to the time when, as a college sophomore, I first discovered cognitive science through his books.

This work would be impossible without a number of incredible collaborators. Shilpa Mody has been the best colleague anyone could hope for. There is no one who provides deeper discussion, sharper insight, or better company. Fiery Cushman took me under his wing when I was first trying to get my bearings. His intellectual and practical contributions to the first paper are tremendous, and he has been a role model as both a scientist and a person ever since.

Many brilliant and supportive people have taught and influenced me over the years. John Vervaeke, who introduced the questions. David Barner, who introduced the methods, and always made his students think hard and dream big. Nathan Winkler-Rhoades, who shepherded my first experience working in developmental research and supported my graduate school ambitions with good advice, kind friendship, and the most delicious food. Mathieu Le Corre, who first gave me the opportunity to do independent graduate research and whose friendship and support have been a joy in all the years since. Noah Goodman, who took the time to meet and talk every week, back when I was a first-year student with barely two decent thoughts to rub together, which I have never forgotten. Lawrence Lessig, who taught me that all of the intelligence and good intentions in the world would be wasted unless they're channeled to good purpose. Stephanie and Rakesh Khurana, along with all of the staff, students and tutors at Cabot House, who have thoughtfully and deliberately created the best possible community to call home.

Being in two labs has the advantage of having twice as many fantastic lab mates. For invaluable discussion, support, and happiness, I am very grateful to Jean-Remy Hochmann, Mariko Moher, Paul Haward, Manizeh Khan, Arin Tuerk, Amy Skerry, Hugh Rabgliati, Tim O'Donnell, Becky Nappa, Mahesh Srinivasan, Eva Wittenberg, Josh Hartshorne, Adena

Schachner, Tanya Levari, Annemarie Kocab, Lindsey Powell, Bria Long, Igor Bascandziev, and Debbie Zaitchik.

I owe thanks to all of the families and all of the anonymous Turkers who participated in this research, as well as the people in the lab who did most of the day-to-day work. Among the most instrumental contributors have been the lab managers: Carlyn Friedberg, Tracy Reuter, Margarita Zeitlin, Ellie Kaplan, Liz Chalmers, Yeshim Iqbal, Rebecca Distefano, Rebecca Zhu, and McCaila Ingold-Smith (first a lab manager, then a friend), Sophia Sanborn (first a friend, then a lab manager and a collaborator). Of course, all the really hard work was done by the most amazing research assistants. Chief among them have been Sophia Clavel, whose contribution cannot be overstated, Maggie Kandel, Karina Tachihara, Amelie Meltzer, Brendon Villalobos, Meltem Yucel, Laura Fitzgerald, Holly Cushman, Lara Bryfonski, Kelly Bell, and Jennifer Kim. Thanks also to the National Sciences and Engineering Research Council of Canada, for the funding that enabled much of this work. No less important are the fine folks at Simon's Coffee Shop, where the vast majority of these pages were written, and whose caffeine is second to none.

None of this would be worth doing without great friends. They have made the journey through graduate school as much about personal as intellectual growth. Thanks always to Dana Gore and to Kate Hobbs, Jennifer Sheehy-Skeffington, Donal Cahill, Phil Enock, Anna Leshinskaya, Alek Chakroff, Emily Bernier, Richard Chang, Steve Frankland, Liuba Papeo, Cristi Proistosescu, Marjorie Freedman, Dana Foarta, Ben Solomon-Schwartz, Cian Power, Bridget Alex, and many others.

Thank you to my family: my wonderful parents, Michael Feiman and Alla Garbar, and to my grandparents for all of their support, encouragement and boundless love. Without their courage and determination, I would not have had a fraction of the opportunities that have led me here. I hope being a doctor of philosophy is close enough.

And finally, thank you to my partner, Molly Baechtold. Her support and love have been the anchor and the harbor.

It's remarkable that humans can make any claims that rise to the level of falsity. Once a thinker achieves this level of clarity and contact with her environment, truth is a mere negation sign away.

-- Paul Pietroski, Framing Event Variables

Rosencrantz: We might as well be dead. Do you think death could possibly be a boat?

Guildenstern: No, no, no... Death is... not. Death isn't. You take my meaning. Death is the ultimate negative. Not-being. You can't not-be on a boat.

Rosencrantz: I've frequently not been on boats.

Guildenstern: No, no, no--what you've been is not on boats.

-- Tom Stoppard, Rosencrantz and Guildenstern are Dead

Chapter 1:
Introduction

1.1 Combinatorial thought

One of the most remarkable things about human cognition is the ability to take a finite number of concepts and combine them to generate an infinite number of meaningful thoughts. You may never have heard the sentence or entertained the thought, “There are no bears on Mars”, but you have no trouble understanding what it means. Not only do you understand it, you can judge that it is very likely true and you can make conclusions on that basis: if there are no bears on Mars, that means there are no brown bears there, no bear cubs, no bears climbing Martian trees. The ease with which you understand new thoughts, make judgments about their truth, and reason through to related thoughts, all have just one plausible explanation – the thinking of complex thoughts is the result of a rule-governed system that combines meaningful component units (words, concepts) in systematic ways.

What are the developmental origins of this system? Is it learned? If so, what kind of evidence would a learner use to induce or create it? Is it most like learning a natural language, or is it actually learning language itself that creates it? Or is this system so fundamental to human thought that it is given innately? What about the individual concepts that can be combined? What are the properties that allow concepts with meanings from very different content domains, like MARS and BEARS, to combine with each other? When a child learns a new concept, is it combinable with others at the outset, or is it possible to learn the content at one time and to gain the ability to combine it with other content later? Does learning a word make its content combinable? Or does natural language only provide a phonological gloss on the combinatorial system that otherwise works independently? Are words just labels, or are they the engine driving combinatorial thought?

These questions constrain each other. The aim of this thesis is to show the outcome of several empirical investigations into these questions and to delineate the mutual constraints. Since the broader project is to triangulate a solution from a few different sides, I do not try to mine a single psychological paradigm or pin down the exact computational details of a specific component of the combinatorial system. Rather, I broaden the empirical target by using a variety of methods and three different populations – infants, children, and adults. The hope is that, although blind men feeling around different parts of an elephant might not come up with a clear picture, they stand a better chance than if they were all groping at the same leg over and over.

I begin with a theoretical discussion that motivates this research program by focusing the many different questions surrounding the nature of combinatorial thought into a single central issue. What makes a mental representation “combinatorial” is not what it refers to or what it represents, but how it combines with other representations; it is a matter of format rather than content. Before describing the three projects that make up the bulk of this dissertation, I examine the most relevant contemporary work in the philosophy of mind and, in its light, consider how empirical methods can be brought to bear on the study of combinatorial representational format. The goal is not to give a comprehensive review, but to make explicit theoretical commitments and to outline a broader framework that can guide empirical investigation into these issues.

1.2 The relationship between format and content

Let me start by making some terms explicit. Following Evans (1982) and Camp (2004), among others, I use ‘propositional thoughts’ to mean thoughts reportable by that-clauses, like the thought that there are no bears on Mars. Propositional thoughts are composed of concepts and have propositions as their content.¹ Both concepts and propositions are types of mental

¹ The mention of a concept as such, as opposed to its referent or denotation, will be indicated by block capitals, e.g. BEAR or MARS. Since I will want to distinguish concepts from non-conceptual representations, I will use italics to

representations. All mental representations have both content and format. The content of a representation is its meaning, while the format is the way in which the meaning is represented.² The same content can be represented in multiple formats, but all formats cannot represent all content with equal fidelity or efficiency, and some formats cannot represent some content at all (for discussion, see Kosslyn, 1980). For example, a tally counting system cannot represent a trillion as efficiently as Arabic numerals, let alone Arabic numerals in scientific notation. And none of these formats can represent the proposition that there are no bears on Mars. Finally, the term ‘conceptual format’ refers to the format that allows representations to represent, or to be combined to represent, propositions.

Testing the format of mental representations poses a problem fundamentally more difficult than testing their content. Experimental stimuli are always content-laden – psychologists are always trying to get participants to represent something about what they are presented with. It is therefore difficult, maybe impossible, to test how representations interact with each other – a property deriving from their format – without the question necessarily being mediated by the particular content of the representations tested. Given that multiple formats can represent the same content, we cannot make conclusive inferences about format from information about content. However, since not all formats are equally good at representing all content, some degree of inference from content to format is possible. If we find that some computational system can represent trillion, it is a good bet that the format of representation is not a basic tally system. What we need are principles that can guide the inference from content to conceptual format.

indicate representations when I am not making any commitment to whether they are conceptual. I will also occasionally use italics for emphasis, and will have to rely on the reader’s charity to distinguish the two uses.

² The present aim is to focus on the format as directly as possible, so I will not make any commitments to how meaning is cached out.

1.2.1 Pure predication

One of the most recent attempts to describe such a principle comes from Tyler Burge (2010a, 2010b), as part of a larger project to distinguish perception from conception. Although Burge is very careful to note that not every non-perceptual representation qualifies as a concept, his insight emerges in the context of distinguishing conceptual from perceptual representation. Burge argues that for a representation to be conceptual, it must involve ‘pure predication’. Here is Burge:

I define ‘purely predicative occurrence’ of an attributive as follows. A purely predicative occurrence of an attributive is one in which the attributive functions predicatively but does not function to make an attribution within the scope of a context-bound, identificational, referential structure to the entity that is (purportedly) identificationally referred to. Such occurrences are purely predicative inasmuch as they function predicatively, but are either outside the scope of a context-bound, identificational referential structure, or are inside such scope but do not themselves function to make an attribution to the entity that the referential application of the relevant context-bound structure functions to refer to. (Burge, 2010a)

On Burge’s proposal, the representation x is F is conceptual if and only if the tokening of F (i.e. the actual creation of the representation F) does not have to depend on the referent of any specific x . It would not be enough to find evidence that an infant can represent mom is happy to attribute to the infant possession of the concept HAPPY, since its representation of happiness in that context is (perhaps inextricably) bound to the perceptual experience of mom.³ Nor would it

³ Although, of course, the concept HAPPY could, if it were available to the infant, be used to predicate happiness of mom within the scope of perceptual reference-fixation. Concepts can live both within and outside the scope of the reference-fixing function. Other representation can live only within the scope of this function. Therefore, according

be enough if the infant could represent mom is happy when mom is not around. Burge is explicit in stating that a distal memory representation of an event triggered in a referential context can be as much a perceptual representation as a proximate, immediate perceptual experience. Rather, representations that count as evidence for the concept HAPPY must be more general or abstract thoughts, entirely independent of any particular referent. This includes thoughts about entire sets, like ALL DOGS ARE HAPPY, or negated thoughts, where the predicate cannot be involved in fixing the referent because it does not apply to it in the first place, as in, MOM IS NOT HAPPY.

Located within the project of distinguishing conception from perception, ‘pure predication’ does not explicitly target conceptual combination at all. It is not clear that Burge is attempting to give an account of combinatorial concepts. Burge’s own explicit goal, however, is less important for present purposes than whether his ‘pure predication’ delineates the same joint in nature as the one I am interested in here. If it does not, it is not necessarily a knock on Burge’s project. But if it does – and it is a plausible candidate – it would be a major finding that a criterion that makes no reference to conceptual combination nevertheless picks out those representations that have this format. One *prima facie* reason to think that the two criteria might pick out the same natural kind is that those representations that Burge takes to be particularly diagnostic of his notion of concepthood – quantifiers like ALL or logical operators like NOT – are paradigmatically combinatorial concepts in the presently relevant sense.

This is an important insight from Burge’s approach, which I will return to. Nevertheless, if taken as a strict criterion for concepthood, ‘pure predication’ runs into problems. On the one hand, it is too strong a requirement. Certain representations may not be extractable from within the scope of reference fixation. When demonstratives like this and that, indexicals like I and you,

to Burge, to diagnose that a representation is a concept, we must have evidence of it being tokened outside of reference fixation.

and referential pronouns like he, she and it are tokened, they are almost always tokened with their meaning dependent on some referent (except perhaps in explicit meta-discourse, like when savvy philosophers are making a use/mention distinction).⁴ But are these not perfectly good concepts that can be combined to form perfectly good propositional thoughts? Take the propositional thought, I AM NOT A BEAR ON MARS and imagine a creature that is capable of representing it, but is only capable of representing the first-person pronoun, I, within the scope of a single referent – itself. This creature would not understand and could not represent that ‘I’ could refer to anything but itself. Nevertheless, it does not seem that anything would prevent this creature from combining that representation, I, with NOT, BEAR, MARS, etc, as long as that representation has the right syntactic and semantic features to be properly combinable. It seems, therefore, that being combinable and being a pure predicate are independent.⁵

1.2.2 The generality constraint

Unlike Burge’s ‘pure predication’, which was not meant to be a criterion for combinatorial representations, the ‘generality constraint’ was proposed by Gareth Evans (1982)

⁴ It may be that Burge does not think that demonstratives and indexicals are ever ‘attributive’ or ‘predicative’, since they are generally the argument of predication, rather than the predicate, in the kinds of examples he gives (e.g. that_i is brown, is the representation brown being predicated on a context-bound, fixed-within-the-scope-of-reference demonstrative). In that case demonstratives could not be ‘pure’ predicates, because they would not be predicates at all. At best, this means that pure predication provides an incomplete criterion of concepthood, since there is a class of non-predicative representations that it does not apply to, and these may or may not be conceptual. More significantly, however, it is easy enough to turn arguments into predicates. Take the perceptually bound representation that_i is I, for a creature referring to itself. I seems perfectly predicative in this case. Being tokened (possibly only ever) within the scope of a reference-fixing cognitive operation, the arguments given above apply as before.

⁵ This is not to deny that there is something to Burge’s intuition. There does seem to be a sense in which the first-person indexical representation I could be a somatosensory representation of the agent’s own body in space, in which case there would be no obvious reason for it to be conceptual in the relevant way. But then contra Burge, it appears that whether I is tokened within the scope of reference-fixation is not a sufficient criterion for determining whether it is a conceptual or perceptual representation. Both a perceptual somatosensory I and a very distinct conceptual combinatorial I might only ever be tokened within the scope of reference-fixation in the same person’s mind. In that case pure predication could not distinguish them.

for just the present purpose. One positive feature it shares with Burge's proposal is that it too can be taken not just as a theoretical desideratum, but also as a methodological guide. Here is Evans:

The thought that John is happy has something in common with the thought that Harry is happy, and ... something in common with the thought that John is sad.... Thus, someone who thinks that John is happy and that Harry is happy exercises on two occasions the conceptual ability which we call 'possessing the concept of happiness'...

We thus see the thought that a is F as lying at the intersection of two series of thoughts: on the one hand, the series of thoughts that a is F, that b is F, that c is F, ..., and, on the other hand, the series of thoughts that a is F, that a is G, that a is H...

If a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception.

The methodological corollary of the generality constraint is that in order to test whether someone – infant or adult – has the concept HAPPY, it is not sufficient to show that the person responds in a way that reflects the content of that concept in a specific case. It is not enough, for example, to know that a baby smiles at a favorite toy, or interprets mom smiling at the toy as evidence that the toy is safe to play with.⁶ The generality constraint requires that the infant be capable of representing that X IS HAPPY for any conceptual X that the infant can represent, and that all the individual predications of HAPPY should be tokens (instances) of the same type.

In the course of empirical research, we can only know whether these constraints are satisfied in relation to each other. We do not know a priori which xs are concepts and which are

⁶ Note the similarity between the requirements imposed by the generality constraint and by pure predication: both insist that representational content is not in itself a good indicator of conceptual format. No coincidence, given the dissociation between content and format discussed at the very beginning. Yet, as I argue below, it is a mistake to think that content can never be a sufficient diagnostic for conceptual format.

not. If it is the case that happy cannot be combined with some representation *z*, is it because happy is not a concept, or because *z* is not one? Just from the evidence that happy and *z* cannot combine, there is no way to tell. The more representations we have evidence of happy combining with, the less likely it is that the different tokens of happy are independent of each other, and the more convinced we become that the representation happy is in fact the concept HAPPY. If, for example, we have evidence that an infant can represent only two *xs* as HAPPY – e.g. herself and mom – it is more plausible that the individual tokens of what looks to the theorist like the concept HAPPY are in fact tokens of two different types (something like me-happy and mom-happy). If, on the other hand, there is evidence that for any new person the infant meets and sees smiling, the infant can represent that this person is happy, it becomes much more likely that the representations of person-a is happy and person-b is happy involve two tokens of the same concept, HAPPY.

The research program that follows from the generality constraint is relatively straightforward – find evidence that for a given representational predicate of interest, it can be attributed to as many individual stimuli as possible. To the extent that there are many such stimuli, and new ones can be arbitrarily introduced, it is likely that the representation is a concept. In this way the generality constraint is very different from Burge's proposal. Whereas Burge is content with only a single *x* combining with *F*, as long as it does so outside the scope of reference, it is relevant for Evans how many *xs* there are and what other representations those *xs* can combine with.

The generality constraint has been extremely influential in the philosophy of mind, but (or perhaps, therefore) it has not gone unchallenged. Critics most often focus on the notion that some representations may be conceptual but are nevertheless not combinable in the way the

constraint demands (see, for example, Peacocke, 1992). One well-known class of examples are sentences that are syntactically well-formed but express semantic category mistakes, so as to be (arguably) uninterpretable. These are sentences like, “Caesar is a prime number” or, “Colorless green ideas sleep furiously”. CAESAR is a perfectly good concept, the argument goes, and so is A PRIME NUMBER, and yet the two cannot be meaningfully combined, so being meaningfully combinable cannot be constitutive of being a concept. Arguments of this sort challenge the generality constraint as a necessary criterion for conceptual thought (but see Camp, 2004, for rebuttal). Setting this issue aside, I focus on whether the generality constraint is sufficient rather than whether it is necessary. If we find that an infant is able to represent x is happy for a variety of x s, does that mean that happy is a concept – that it is freely combinable in a system of thought with other concepts?

Take any cognitive mechanism that functions to identify a particular stimulus type in the environment. This can be a basic perceptual mechanism – say, one that represents that an object is blue or round – or a mechanism that represents more abstract features, such as the ‘input analyzers’ that have been proposed for representing objects, agents, and approximate magnitude (see Carey, 2009). Any such mechanism will be able to represent that x is F , where x is a variable ranging over some set of stimuli and F is the predicate (e.g. the property, category, or attribute) that this mechanism processes and represents. If F is the representation happy and the various x s are all the individuals that the infant can represent as being happy, then the generality constraint looks to be met to the extent that there is some significant number of x s that can meaningfully be represented as being happy.⁷ Note that it is not necessary here that the mechanism for representing F be dedicated to representing F and nothing else, or that it be modular, innate, etc.

⁷ Although Evans would want the criterion to be predication over all x s that are concepts, I say “a significant number of x s”, because we have no independent way of knowing which x s are concepts and which are not.

It just needs to be capable of assigning some represented property F to some fairly wide set of stimuli (following Evans, to all of the xs to which other properties – G, H, etc. – can also be assigned). There is only one important requirement on F – that it be tokened specifically as a predicate bound to different arguments. It would not be enough for an infant to see many different happy people and represent happiness every time. To meet the generality constraint, it is important that the infant actually represent person-A is happy as distinct from person-B is happy.

But if this is what it takes for a representation to qualify as a concept, it is far too weak a requirement. Any cognitive mechanism that represents any stimulus in the world in a way that involves recognition or categorization of a property is going to abstract away from some properties of the individual token stimuli that trigger that representation. A mechanism that responded only to one token – to one specific instance of a stimulus – or only to stimuli identical in every way to the same token, would be very nearly useless. A mechanism that represented the abstract property without binding it to a particular object would hardly be better. It would not allow for any behavioral (or even further cognitive) response directed at the entity to which the property applies, would not be able to distinguish between entities that have the property and ones that don't, would not in fact be able to identify the source of its happiness (or whatever other representation).

Imagine a study in which infants see two rapidly flashing video feeds. To one side of the seated infant, one feed always flashes the exact same picture on-and-off. The picture contains three colored squares, and nothing ever changes. The other feed, on the other side of the infant, rapidly flashes different pictures, also of three colored squares. But in this feed, two of the squares swap colors every time the new picture flashes. What if infants, by 7.5 months of age,

differentiate the two streams, attending longer to the feed where two squares are swapping colors (as in fact they do, see Oakes, Ross-Sheehy & Luck, 2006)? To pass this task, infants cannot simply be representing blue and red, unconnected to any other representation. They must be binding color features to objects to be able to notice that the object's color changes. This experiment shows that infants represent x is blue for at least two x s, and do indeed bind these objects to the predicate blue. It shows the important result that infants can solve a version of the 'binding problem' of vision in short term memory, binding different perceptual features together – in this case, objects to colors. But it does not come close to showing that infants can combine their representations of blue and red productively to form propositional thoughts. It does not show that infants' representation blue is the concept BLUE.⁸

The generality constraint imposes a condition on mental representations having predicate-argument structure from the theorist's-eye view. But that alone is not a sufficiently strong constraint to identify conceptual combinatorial thought. Despite its influential history in the contemporary philosophy of mind, at the very least, some additional criterion is needed to distinguish concepts from other types of representations.⁹

⁸ It is no help that Evans requires all of the X s combinable with (represented as having the property) blue to also be combinable with other predicates. Imagine, as in fact seems to be the case, that the same infant cognitive system contains other input analyzers – ones that recognize objects and estimate approximate relative size, for example (Carey, 2009). All of the same blue objects could serve as perfectly good input to these processors, triggering the representation that they are all objects, and that their approximate common size is some specific quantity, yea big. But just in the case of the blue representation, the object and approximately-yea-big representations do not bring this network any closer to conceptual combination. Each input analyzer, and each representation produced by it, is independent of the others and none need be combinable together. Being able to recognize that some specific toy is blue and, separately, that it is an object (and thus ought to remain solid and not to teleport), is not the same as being able to represent the proposition that it is a BLUE OBJECT. Representational co-reference is not conceptual combination.

⁹ Another potential objection to the generality constraint is that it could characterize a very small-scale, encapsulated combinatorial system, such as the "language" of bee dance or bird vocalizations, which, despite having some combinatorial structure, theorists may not want to call a 'conceptual' system. I do not pursue this objection here because it goes right to the question of just how much systematic combinability is needed to grant a creature the possession of a conceptual combinatorial system. This debate seems to me to be a matter of both terminological taste and preferred degree of species-chauvinism. The critique above is stronger because it argues that the generality constraint is not a sufficient condition no matter how many different predicates can be combined with how many different arguments.

1.3 The way forward

It appears that both the generality constraint and the requirement of pure predication are not sufficient, and may not even be necessary, as criteria for concepthood. The approach that I propose to take here nevertheless draws on both of these principles. But while the goal for both Burge and Evans is to come up with a criterion or test which can be applied to any representation to tell whether it is a concept, the approach on offer here is a bit more modest. I propose to look for evidence of specific representations that can only be conceptual, and once these are found, use them to examine which other representations they can combine with. If a concept can combine with a representation, that other representation must itself be a concept. In this way, a single concept can be used as a wedge into the rest of the conceptual network. But which concepts can serve this purpose? Burge hit on the right examples, though he used them for different reasons.

The key is this: representations of abstract logical operators, like not and all, are the rare case where content is diagnostic of format. The content – the meaning – of these representations just is in their combinatorial properties. That is not to imply that other concepts somehow do not have comparable combinatorial properties – they must, or else they would not be concepts in the present sense. But while the content of bears can be represented in either a conceptual or non-conceptual format, logical representations can only be concepts because they have nothing but combinatorial properties. Not is a unary function that takes as input one truth-value (True or False) and produces the other truth-value as output. Any creature able to represent not X, even just for one solitary x (perhaps because they only have one x that is in the right format), and even just if that one x is tokened in a referentially-dependent way, possesses an ability to combine concepts systematically to form propositional thoughts. If the representation not really has a

truth-functional meaning that takes truth-values as inputs and produces truth-values as outputs, then it is the concept NOT. And if the representation x can combine with this conceptual NOT, then x has the combinatorial properties that make it a concept, too. X and NOT may not be combinable with many other representations if those other representations are not yet conceptual, but they would already be components of a nascent system of combinatorial propositional thought.¹⁰

In this way the presence of the concept NOT can itself serve as a diagnostic tool for identifying whether other representations are concepts. This approach provides a way to resolve the epistemic problem that was raised earlier in the discussion of the generality constraint. If we establish that a person is capable of representing both happy and z, but cannot combine these to form what would be the propositional thought that Z IS HAPPY, how do we know if only one or both of the representations are not conceptual, and therefore not combinable? Unlike happy, NOT is a representation that cannot be anything but conceptual. If we know that a person is in possession of NOT, and yet still find that the same person cannot represent NOT Z, then we know that the fault must lie with the representation z – that it is not a concept. Representations of logical operators allow us to grab hold of one of the horns of the disjunction, and the interpretational ambiguity that poses a difficulty for the generality constraint transforms into an investigative method, where we test concepts one by one to see if they can combine with NOT and with other logical operators.

There is still a significant challenge, however. For any task that requires the infant to negate some specific content – whatever the content in the experimental stimuli – there can be

¹⁰ The picture here is one of a system of thought that gets filled out with concepts as the developing person acquires them. For the purpose of characterizing this system, it is then possible to stay agnostic about how individual concepts are acquired -- whether they are acquired through language, explicit education or training, a child's exploratory hypothesis formation and testing, or some interaction between these and still other factors.

domain-specific representational imposters. Within any particular content domain or even within a single computational circuit, nothing prevents there being a less general representation, say not-blue, which performs only the single limited function of representing that some stimulus is not blue, but that is entirely separate from a logical conceptual NOT. In fact, the finding that even neonates can habituate to some repeated stimulus and dishabituate when that stimulus changes (for example, Brody, Zelazo & Chaika, 1984; Bartoshuk, 1962) means that, at some level of description, there is a computation that uses a kind of negation, since they recognize that the new stimulus is not the old one. Even the sea slug *Aplysia californica*, whose cognitive architecture is implemented in a comparatively simple nervous system of only 20,000 neurons, can habituate when stimuli are presented repeatedly and dishabituate when they change (Pinsker, Kupferman, Castellucci, & Kandel, 1970). Logic gates – physical devices set up to implement the input-output mappings specified by logical operators – are among the most fundamental building blocks of all computation. A NOT gate is an exceedingly simple thing – it is just a circuit that blocks or inverts the signal it receives, taking 1 and outputting 0; taking 0 and outputting 1. NOT gates and other logic gates can be trivially implemented in neural networks, in the connections between individual neurons, in the action potential signal of a single neuron and even in intracellular molecular structures (see Gallistel & King, 2009 for discussion). The difference between a NOT logic gate and the concept NOT is, at least in part, in the range of its application. A single NOT-gate is wired to take input from a single source, deep within a computational system. For example, the biochemistry of a single neuron could function to implement a NOT-gate so as to produce a habituation response to the kind of stimulus that the neuron is typically responsive to. In contrast, the concept NOT, since it is part of an entire combinatorial system,

needs to operate as a function that can be called and reused with many different arguments – many different concepts.

If telling whether a given representation is the concept NOT or a much more input-limited NOT-gate is a matter of seeing how many other representations the representation of not can combine with, are we not just back to Evans' generality constraint? Not quite. The argument given previously still holds – we ought not grant possession of the concept BLUE to a creature on the basis of its being able to predicate blue of a variety of arguments. But the earlier point that logical operators are different holds as well. The content of logical operators is inherently combinatorial, and combinability with a wide range of inputs is precisely the thing that distinguishes an encapsulated and computation-internal NOT-gate from the conceptual operator NOT. Only with a logical operator like negation does this generality ensure that we are dealing with a concept – a part of a combinatorial system of meaningful thought – and not an input-analyzer that can process a variety of different inputs. Put another way, while the generality constraint alone is not sufficient to diagnose a concept, the generality constraint together with the content of a representation being necessarily semantically combinatorial, are jointly sufficient.¹¹

The framework laid out above is meant to make explicit theoretical commitments, as well as to give the overarching theme for the specific projects presented below. The rest of this chapter will introduce the three papers that comprise this dissertation. It will relate each paper to

¹¹ There is another important difference between the generality constraint as it is used here and by Evans (1982). The generality constraint here is diagnostic (sufficient) rather than constitutive (necessary). A creature could in principle be in possession of conceptual NOT, but not of any other concept. In that case, NOT would not be any less conceptual for its inability to combine with other representations. The fault would lie with the paucity of other concepts, even if there would not necessarily be any way to tell that the creature does possess NOT in this case. Importantly, the use of the generality constraint as diagnostic in the case of logical operators in particular, rather than constitutive for all concepts in general, also means that seeing whether a logical representation satisfies the generality constraint is only one possible research strategy, and not the only one.

a particular research strategy that addresses some of the central questions concerning combinatorial thought.

1.4 The Three Papers

1.4.1 The first paper

In answering the question of whether a conceptual system of thought is innate or in some way learned or acquired, we are interested not only in evidence of the presence of a concept but also in evidence of its absence. While evidence that a creature non-linguistically represents not- x for some specific x is ambiguous between possession of conceptual NOT and some domain-specific imposter not- x , what about evidence that the creature fails to represent the negation of x in either format? At the very least, such a finding means the lack (or failure of deployment) of some limited not- x representation for that x . It does not, however, necessarily mean that the creature is not in possession of conceptual NOT. The reason should be clear given the preceding discussion -- it could be that the x involved is non-conceptual, even if the creature can represent NOT in combination with some other X s.

In the first paper, we test infants' understanding of two basic types of goals – approaches and avoidances. We use an established paradigm for testing infants' representations of goal-directed action (Woodward, 1998). The goal-directed action here is a hand appearing from behind a screen, touching and picking up one of two objects on stage. In the classic Woodward paradigm, two objects (e.g. a ball and a bear) are presented and a hand is seen repeatedly grasping one of them. At test, the objects change locations and the hand either grasps the same object as before in a new location or a never-before-grasped object in the old location. In the original study and in many subsequent replications infants have been found to dishabituate if the hand reaches for a new object but not for the old object in a new location.

We modify the paradigm to create two distinct conditions – one ‘Approach condition’, where one object is always approached (while, one by one, an array of others are avoided in relation to it), and another ‘Avoidance condition’, where one object is avoided (in contrast to many others, which are approached, one by one). We then see whether infants look longer to an event that violates the previous pattern – finally reaching for a new object in the Approach condition or finally reaching for the previously avoided object in the Avoidance condition. Given that infants succeed in representing a consistent approach in the Approach condition, they could succeed in the Avoidance condition in one of two ways. They could have a distinct representation, avoid, or they could combine the approach representation we know them to have with a conceptual NOT. Finding that they fail to notice the violation event in the Avoidance condition would be evidence that they either do not have access to conceptual negation or else that their representation approach is not the concept APPROACH, and therefore cannot be combined with negation.

1.4.2. The second paper

Despite the abstract meaning of logical negation, the word “no” is among the earliest words children produce. This might suggest that children have the logical meaning of the word very early, perhaps before they begin to learn language at all, and that mapping the word to the logical concept is relatively easy. Except that it is not clear that children produce “no” with a logical meaning as soon as they produce the word. Early on they seem to use it primarily to reject parental commands, offers or prohibitions. However, data from production alone is ambiguous; children’s use of “no” to reject could reflect either their having mapped the word only to some representation like get-away-from-me, or it could reflect their using the same logical meaning as adults for a narrower set of speech acts. After all, adults who have a logical

negation meaning for “no” can very reasonably use it for rejection, to express the propositional thought, I DO NOT WANT THAT. Maybe children simply have more reason or occasion to make rejection speech acts using the same truth-functional concept than do adults.

In the second paper, we investigate the relationship between learning the words that express logical negation and acquiring that concept. Do children have a concept of truth-functional negation before they learn the words “no” and “not”, so that learning these words is a matter of mapping a label to a preexisting concept? Or does learning the word precede learning the concept? If the latter, is it because children do not have the concept at the time they first learn the word, or because mapping the concept to the word is a difficult language-learning problem? To make progress on these questions, the second paper compares two aspects of the linguistic development of negation: the developmental courses of the production of the words “no” and “not” and the comprehension of the logical meaning of those words.

The strategy taken in the second paper is to look for a gap between comprehension and production for both words. While “no” is one of the earliest words children produce, they do not begin to say “not” until early in the third year. If we find that for each word, either logical comprehension and production track together or comprehension precedes production, it would suggest that the child has the concept before the word and that mapping the word to the concept is constrained by linguistic properties of each word, such as their input frequencies and their different grammatical properties. If, on the other hand, production precedes comprehension, there are two options to explore. One is that the child does not yet have the adult-like concept at the time when the word is first learned. This is the case in which learning the word could possibly play some causal role in the acquisition of the concept. The other possibility is that the adult concept is present, but that mapping the word to that concept is itself a hard problem

because the hypothesis space of possible mappings is initially insufficiently constrained for the child to identify the right one (see Snedeker and Gleitman, 2004).

To test the comprehension of “not”, we present children with a situation where a ball is hidden behind an occluder in either a bucket or a truck, and the child is told verbally that the ball is “not in the bucket/truck”. On control affirmative trials, children are told where the ball actually is (e.g. “It’s in the bucket/truck”). When testing children’s comprehension of logical “no”, we modify the paradigm to include a question-answer dialogue with the child’s parent. The parent asks, “Is it in the bucket/truck?” And the experimenter answers either, “Yes, it is” or “No, it’s not”. We find that children’s comprehension of both “no” and “not” emerges at the same time, around 24 months, and patterns with their production of “not”. This is about a year later than when they begin to produce “no”, suggesting that “no” has a different meaning for them for the first year that they are producing it, and is retrofitted to map to the concept of logical negation only later. In a control experiment, we test the deflationary account that children’s improving performance between 19 and 28 months is due to an increasing ability to inhibit attention and search find that when children are shown that the bucket is empty.

The finding of a year-long comprehension-production gap for “no” suggests either that children do not have the concept of truth-functional negation before 24 months – the age at which they begin to succeed on the comprehension task – or else, if they do have the concept, that mapping the word to the concept poses a particularly difficult language-learning problem. That “no” and “not” are comprehended at the same time constraints the possible nature of a language-learning difficulty. The difficulty cannot be with the specific grammatical or frequency-based features of “no” and “not”, since all these are quite different from each other. If it is a language learning problem, it is likely to be the same problem for both words. One

possibility is that to map either word to a logical negation concept requires knowing a significant amount of the syntax and lexicon of the language. It is hard to understand “There are no bears on Mars” unless one understands the meaning of “bears” and “Mars” and how they combine. Perhaps children need to know enough of the language to figure out the semantic role both “no” and “not” play as logical negation. Once they have acquired enough linguistic knowledge, they might make the mapping for both words with equal ease. We discuss how these two possibilities might be addressed by future work looking at a special population that provides a natural experiment that separates conceptual sophistication from linguistic competence – internationally adopted toddlers who begin to learn English much older than native learners.

1.4.3 The third paper

The third paper is different from the other two in a number of ways. The participants tested in it are adults. Rather than trying to tell whether they have access to a logical concept like NOT or ALL, this paper looks at the structure of logical combinatorial concepts in their fully developed states. We examine quantifier words like “Every”, “Each”, and “All” and ask what structure these have in combinatorial language. The fundamental question in this paper is: in the process of assembling complex meanings from the combinations of individual words, which properties of meaning are the combining operations sensitive to and which ones do they abstract from? On one level of description, all three of the quantifier words above instantiate the same logical operator (the universal quantifier). However, fine-grained patterns of grammaticality judgments also show meaning-relevant differences between all three quantifiers. Using a priming paradigm, we ask whether interpreting one ambiguous sentence with one of these quantifiers affects how participants interpret another similar ambiguous sentence with the same or another quantifier. We compare the presence and strength of priming between sentences with different

universal quantifiers, as well as between sentences with one universal quantifier and another quantifier with a different meaning (e.g. “Three”).

The priming methodology takes advantage of a linguistic phenomenon called scope ambiguity, which gets at the heart of how combinatorial meaning is assembled. Sentences with two different types of quantifier words (e.g. a universal like “Every” and an existential like “A”) have a systematic ambiguity in the possible interpretations they can have, having to do with which quantifier takes scope over the other. For example, in the sentence “Every boy climbed a tree”, one meaning has the universal taking wide scope over the existential – for every boy, there is a potentially different tree – while the other meaning has the existential taking wide scope – there is a tree, which every boy climbed.

Using a picture-choice paradigm where both alternatives are presented and participants are asked to match one of the pictures with the ambiguous sentence, the first experiment in this study provides baselines for preferences between the two readings for many combinations of different quantifier words in subject position with the quantifier “A” in object position (e.g. “Every boy climbed a tree”, “Each boy climbed a tree”, etc.) This baseline experiment shows that the preferred reading depends strongly on the subject quantifier word used. From previous work on scope ambiguities, we know that the preference for one reading over another can be primed by a preceding ambiguous sentence with different noun content but the same two quantifiers generating the same kind of ambiguity (eg. “Every hiker climbed a hill”), where one or the other type of reading is forced (eg. participants must choose the picture where many hikers are on the same hill; Raffray and Pickering, 2010). We adapt the priming paradigm to look at priming effects across different quantifiers, testing all permutations of “Every”, “Each”, “All” and number quantifiers in both prime and target position.

We find a priming effect across all prime-target pairs with the same quantifier words (i.e. from “Every-A” to “Every-A”, “Each-A” to “Each-A”, “All-A” to “All-A” and “Three-A” to “Three-A” sentences), but find no effect across different quantifiers (eg. from an “Every-A” prime to an “Each-A” target). Further experiments show that the priming effect does not depend on overall lexical similarity between prime and target, since the priming effect obtains when prime and target sentences have different verbs as well as different nouns, and when they have different numeral quantifier words (eg. there is a priming effect between “Three boys climbed a tree” and “Five hikers climbed a hill”). This last experiment suggests that the combinatorial machinery of meaning is blind to some underlying semantic differences – like those between the meanings of “four” and “five” – but not others, like the differences between the universal quantifiers.

This last finding is worth emphasizing for its implications both for the structure of natural language semantics and the interface of language with thought. Setting aside several implications for specific classes of linguistic theories (discussed in more detail in the paper), one key result is the division of labor between the combinatorial semantics of words and the concepts those words point to. On anyone’s theory of semantics, the combinatorial properties of “Three”, “Four” and “Five” are the same; the only difference between them is in the cardinality concepts they point to. That the priming effect is specific to each individual universal but is not sensitive to the differences between different numbers provides empirical support for this foundational theoretical assumption.

1.5. The end of the beginning

The papers that comprise this dissertation are concerned with the format of combinatorial thought rather than with the content of any specific representation. To study format requires

finding properties of representations that abstract away from what the representations are about and looks at what they can do in terms of interacting with other representations. What is the format of representations involved in a system of combinatorial thought, which is capable of recombining a finite array of concepts to represent an infinite array of meaningful propositions? I have argued against two popular criteria for identifying this conceptual format: Burge's pure predication and Evans' generality constraint. I argue that it is nevertheless possible to study the emergence and structure of conceptual format empirically by focusing on just those representations that have combinatorial properties – and nothing but combinatorial properties – by virtue of their content, then testing other target representations for their ability to combine with these concepts as a test of the targets' conceptual format. I now turn to describing the collaborative work wherein the rubber meets the road, empirically investigating the emergence of logical negation in infants and children and the structure of natural language quantification in adults.

Chapter 2:

Infants' representations of others' goals: Representing approach over avoidance

2.1. Introduction

Convergent research indicates that young infants, even neonates, create representations of agents and attribute intentions to their actions (e.g. Senju & Csibra, 2008; Sommerville, Woodward & Needham 2005; Woodward, 1998; Luo & Johnson, 2009; Gergeley & Csibra, 2003; Onishi & Baillargeon, 2005; Senju, Southgate, Snape, Leonard, & Csibra, 2011; see Baillargeon, Scott, He, Sloane, Setoh, Jin, Wu, & Bian, 2014 for a review). Much less is understood, however, about the form those representations take and how they are computed.

Consider a paradigm introduced by Woodward (1998) to investigate infants' representations of goals. In this paradigm infants are habituated to an experimenter repeatedly reaching for and touching one of two objects (e.g. a ball over a bear). On the critical test trials that follow, the two objects switch locations and the experimenter reaches again, either for the same target (the ball contacted during habituation, which is now in a new location) or for the same location (the bear that the experimenter had never before touched, now sitting in the location where the experimenter had formerly reached). Infants dishabituate to a reach to the new object in the old location, but not to a reach for the old object in a new location. This paradigm has been extended to displays where the agent picks up the object (Phillips, Wellman & Spelke, 2002; Sodian & Thoermer, 2004), and ones where the entire agent approaches the object rather than reaching for it (Hernik & Southgate, 2012; Lakusta, Wagner, O'Hearn, & Landau 2007). In all of these cases, experimenters concluded that infants' intentional construal of the event was of the agent fulfilling its goal to contact the object. More recently, these and other manipulations have also been interpreted as providing information to the child that the agent has a positive disposition toward the approached object, which in turn leads to the prediction that the agent will approach the object in the future (Baillargeon, et al., 2014).

There is, however, another possibility. When an agent consistently chooses a ball over a bear, this action is consistent not only with the agent having a goal to obtain the ball, but also with the agent not having a goal to obtain the bear, or with a goal to avoid it. Either of these representations alone would be sufficient to explain dishabituation in the Woodward paradigm, and neither is mutually exclusive with the standard interpretation: Infants might represent goals alongside non-goals, or approach alongside avoidance. This ambiguity is present in the account where infants represent agents as having a particular disposition towards an object as well. Not only is the Woodward paradigm ambiguous with respect to the evidence the child uses to establish the agent's disposition (approach to one object or avoidance of the other), it is also ambiguous with respect to which attributed disposition underlies the child's attention to the unexpected event (attributing to the agent a positive disposition toward the approached object or a negative disposition toward the avoided object).

Although these alternative possibilities have not received much attention, a number of considerations lend them some plausibility. From an evolutionary perspective, avoiding is often more critical for survival than approaching –a single encounter with a predator could well be deadly. Representing, noticing and learning from others' avoidance goals are therefore likely to be important for young humans, as for other animals. Furthermore, recent work on a negativity bias in both adults' and infants' processing of valenced information suggests infants do attend to negative information. Three-month-old infants prefer neutral over antisocial agents, but not prosocial over neutral agents (Hamlin, Wynn, & Bloom, 2010). Studies of social referencing show that infants generally modify their own behavior more in response to negative than to positive affective information from their caregivers (see Vaish, Grossman, & Woodward, 2008, for a review). Twelve-month-old infants faced with an ambiguous new toy play with it less if

their caregiver looks disgusted rather than neutral, but do not play with it more if the caregiver emotes positively, rather than neutrally, towards the toy (Hornik, Risenhoover, & Gunnar, 1987).

Additionally, a number of other studies test infants' sensitivity to others' positive and negative emotions and preferences, by providing both kinds of information within the same condition. These studies indicate infants' sensitivity to valenced intentional information, even if they do not allow us to compare positive to negative directly. Thus, by 18 months, infants will give an agent an object she emoted positively rather than negatively towards, (Egyed, Király, & Gergely, 2013), match the food preference of a prosocial or novel agent who indicated liking one food and disliking another (but not of an antisocial agent; Hamlin & Wynn, 2012), and override their own preference to give an agent a food that the agent has shown a preference rather than a dispreference for (Repacholi & Gopnik, 1997). This literature suggests that infants in the Woodward paradigm might indeed attend to the consistent avoidance, and perhaps even attend to it preferentially over a consistent approach, and might indeed attribute a negative disposition toward that object to the agent.

Most studies using the Woodward paradigm are ambiguous on this point, because every trial with a persistently reached-for object has always included a persistently not-reached for object. Some relevant evidence comes from studies where the habituation display involves only one object that is consistently approached. If an agent simply approaches a single object along a straight path, as in the Woodward paradigm, infants do not expect the agent to continue approaching that object (Luo and Baillargeon, 2005). However, when the agent approaches the object by taking an efficient path around an obstacle, infants successfully predict an approach to the same object during test (e.g. Biro, Verschoor, & Coenen, 2011; Hernik & Southgate, 2012). These studies indicate that infants can represent a consistent approach, since there is no

consistent avoidance during habituation. However, they do not bear on the question of whether infants also represent avoidance given evidence consistent with both avoidance and approach, as in the canonical and widely-used Woodward paradigm.

The present study addresses this ambiguity through a modification of the original Woodward paradigm. As in the original, two objects are present during habituation. But, while one of the objects stays the same across all trials, the other object's identity changes on every trial. Infants see one of two habituation displays: either a consistent reach to the same fixed object (the Approach condition), or a consistent reach to the always-novel, variable object and, therefore, a consistent avoidance of the fixed object (the Avoidance condition). The Approach and Avoidance conditions each provide equivalent evidence for an approach goal or an avoidance goal, as well as for a positive or a negative disposition toward the fixed object, respectively.

If infants only require that one fixed reached-for object be paired with a foil in order to establish that the agent has the goal of contacting that fixed object, then they should succeed at the Approach condition. Similarly, if infants need only one unreached-for fixed object and a foil to establish that the agent has the goal of avoiding or of not picking up the fixed object, they should succeed at the Avoidance condition. If infants succeed in both of these conditions, that would suggest that imputing positive and negative valences, goals to approach and goals to avoid, are equally available to young infants as they make sense of the events in the basic Woodward paradigm. We begin our investigation with 7-month-old infants, who have been shown to succeed robustly in multiple versions of the basic Woodward paradigm (Woodward, 1998; Luo & Baillargeon, 2005; Luo & Johnson, 2009).

2.2. Experiment 1

2.2.1. Method

2.2.1.1. Participants

Sixteen full-term infants participated in the Approach condition (mean age 7;3, range 6;15-7;15, 11 female) and another 16 in the Avoidance condition (mean age 7;1, range 6;15-7;15, 10 female). One other infant was excluded from analysis and subsequently replaced due to fussiness, and another one due to parental interference.

2.2.1.2. Stimuli

Up to 19 objects were used for each infant: one “fixed” object present on every trial and 18 “variable” objects each used on one trial only. The identity of the fixed object was varied across participants. The objects were similarly sized, but differed in material, texture, color, shape and semantic category. They included, for example, a small section of a wooden branch, a rubber duck, a plastic carrot, a shoe brush, and a jewelry box.

2.2.1.3. Setup

Infants sat on a parent’s right knee about 90 centimeters from a stage surrounded by black curtains. Parents were instructed to close their eyes throughout the experiment and not to direct the infant’s attention. The stage had exits on each side with black curtains shielding the experimenters from view. A black plank spanned the width of the stage, extending out of both side exits. This plank had strips of velcro attached, corresponding to the three locations where objects could be placed. Objects, each with a strip of velcro underneath to keep them from rolling off or from wiggling relative to the plank as it was moved, were placed on the sliding plank and shuttled to one side or the other such that the fixed object and one variable object were in view on each trial (see Fig. 2.1).

A camera behind the infant recorded events on stage, while another hidden in the stage curtains recorded the infant's gaze. Two trained coders in another room (hence blind to condition) recorded infants' looking times. A primary coder's input determined the end of a trial. Average percent agreement on all trials was 94.1 for Experiment 1, reflecting the proportion of 100 ms time segments during which both coders agreed about whether the baby was looking at the display.

2.2.1.4. Procedure

Both conditions consisted of a habituation phase followed by a test phase. Habituation lasted until the infant's looking time on three consecutive trials was less than half that for the first three trials, with a maximum of 14 habituation trials. This habituation phase was followed by four test trials. In the first and third test trials the hand reached to the unexpected object, while the second and fourth trials maintained the pattern of reaching established during habituation. All trials began when the infant looked onstage for at least 0.5 seconds and ended when the infant looked offstage for at least 2 seconds.

In the Approach condition, all habituation trials consisted of a hand emerging from the back stage curtain, reaching for the fixed object, picking it up and shaking it in the air until the end of the trial before replacing it and retreating behind the curtain. Avoidance condition habituation trials consisted of an identical action performed on the variable object. No curtain or barrier was used to block view of the stage between trials.

The variable object was swapped off-stage between each trial. The fixed object sat at the center of the plank and remained in full view of the infant throughout the experiment. Variable objects were placed on either side of the fixed object and shuttled on- and off-stage on the plank. When a new variable object was shuttled onstage from the right, the old variable object exited to

the left, and vice-versa. Once off-stage the old variable object was replaced with a new one.

Thus, the relative location of the fixed and variable object (right stage versus left stage) alternated on every habituation trial, as did the direction of the hand's reach (see Fig. 2.1).

In both conditions, the first test trial was always an unexpected trial, in which the hand reached for the variable object in the Approach condition and the fixed object in the Avoidance condition. This broke the left-right alternation of reaches. However, because the pattern of alternation was established identically and broken identically across both conditions, differential patterns of dishabituation between them cannot be attributed to the sequence of right stage versus left stage reaches.

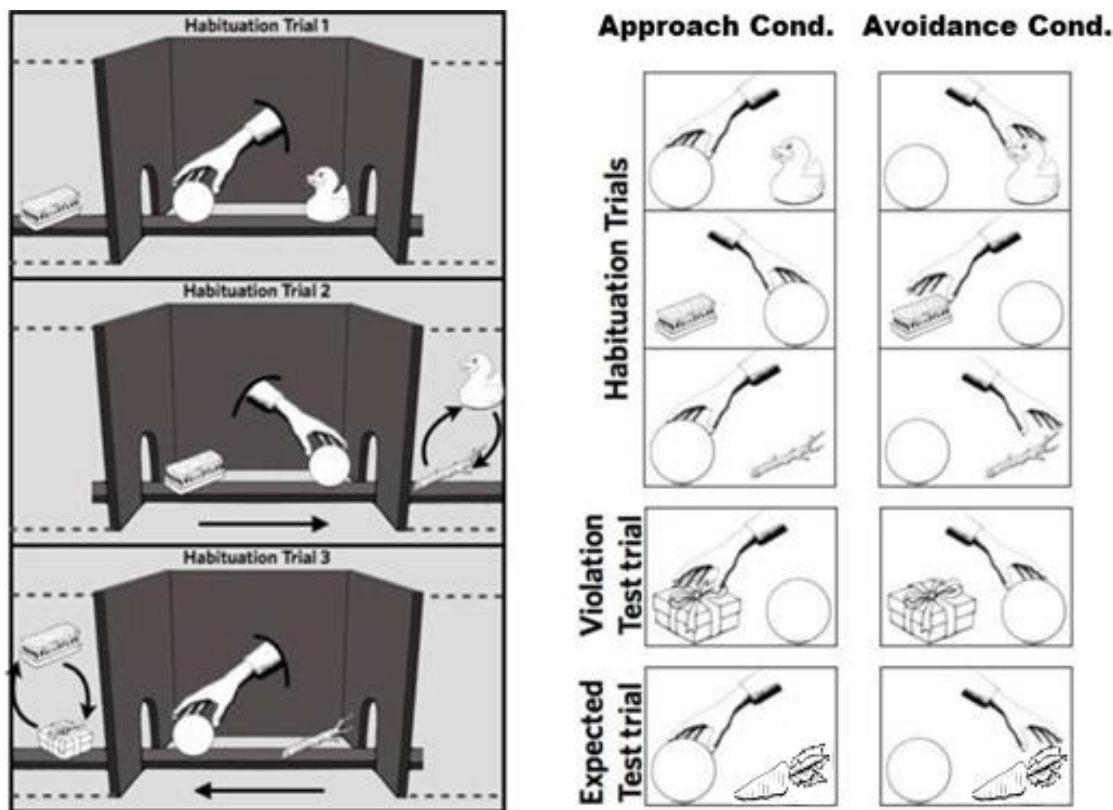


Fig. 2.1: The stage setup (left) allows for a single fixed object to be shuttled from one side of the stage to the other via the sliding bar on which it is placed. The variable objects are switched after they are shuttled off stage and out of sight of the infant. The procedure, illustrating the pattern of alternating reaches (right) is shown only for the Approach condition, although all conditions involved the same alternation of positions for the fixed and variable objects. In the Approach

condition, the hand always reaches for the ball during habituation. On the first test trial, it reaches for a novel object (and breaks the alternation) for the first time. This first test trial always violated expectation.

2.2.3. Results

2.2.3.1. Habituation Trials

The first two panels of Figure 2.2 shows the looking times during the first three habituation trials, the last three, plus the violation and expected test trials of the Approach and Avoidance conditions of Experiment 1. Infants habituated similarly in both conditions. Thirteen reached criterion in the Approach condition and 14 in the Avoidance condition. An ANOVA examined the effects of condition (Action vs. Avoidance) and habituation trial block (first 3 vs. last 3) on looking times during habituation. There was a main effect of habituation block, $F(1,30) = 52.59$, $p < 0.0001$, indicating that infants looked less over time (M first 3 = 19.52 seconds; M last 3 = 8.78 seconds. There was no main effect of condition, $F(1,30) = 0.2$, $p = 0.66$, nor did condition interact with habituation block, $F(1,30) = 0.95$, $p = 0.34$. Thus, infants entered the test trial phase in similar attentional states.

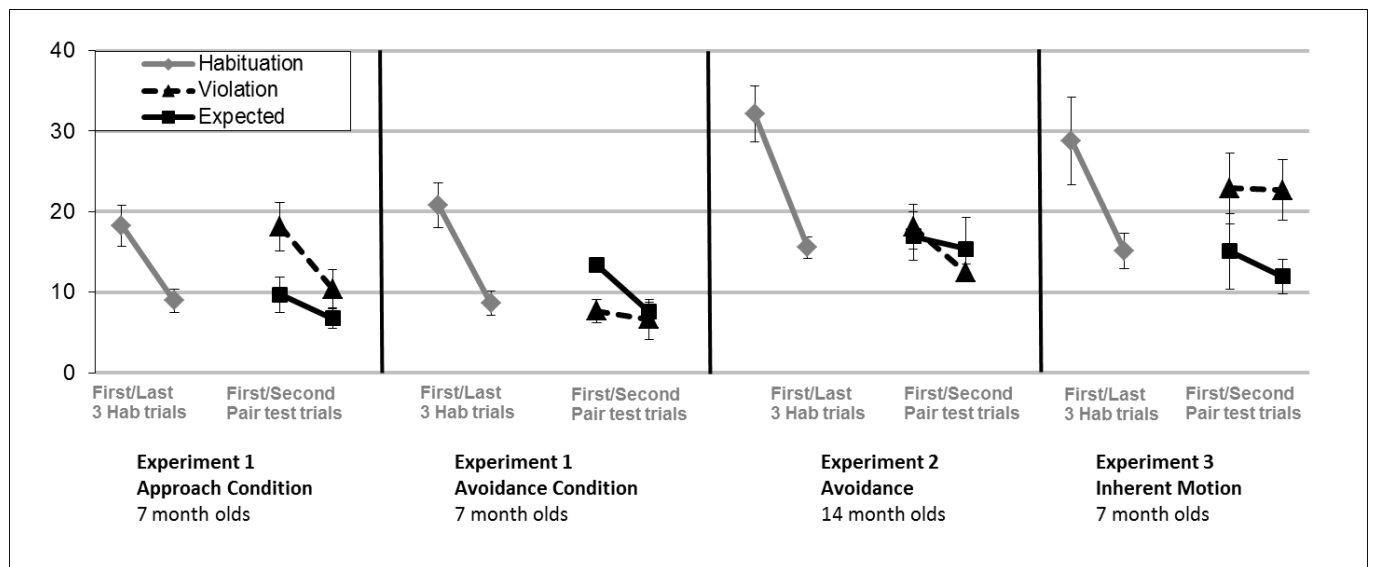


Fig. 2.2: Results from both conditions of Experiment 1 and Experiments 2 and 3, showing looking time per trial averaged across participants. Error bars indicate 1 S.E. from the mean.

2.2.3.2. Test Trials

We excluded data when the difference between the violation and expected test trials exceeded 2 standard deviations from the mean difference across subjects, considering the first and second pairs of test trials separately. One pair of second test trials was excluded in each condition.

An analysis of variance examined the effects of the within-subject variables of trial type (Violation vs. Expected) and trial pair (first vs. second pair) and the between-subject variable of condition (Approach vs. Avoidance) on looking times during the test trials. Critically, this revealed an interaction between trial type and condition, $F(1, 28) = 19.81, p < 0.001$, indicating dishabituation to violation trials selectively in the Approach condition. This pattern was strongest for the first pair of test trials, leading to a significant three-way interaction between trial type, trial pair and condition $F(1, 28) = 4.69, p = 0.04$. Finally, a main effect of trial pair $F(1, 28) = 10.85, p < 0.01$ indicated that looking times were longer overall on the first pair of test trials. There were no other significant main effects or interactions.

Planned comparisons illuminate the source of the interaction between trial type and condition. The results of these comparisons and the associated parametric statistics are shown on Table 2.1.¹²

¹² Order of expected/unexpected trials in violation of expectancy experiments are usually counterbalanced, because usually they have some salient difference in overall structure that in itself might increase looking time. For example, in the Woodward paradigm, the location of the two objects is switched between habituation and test trials. However, in our design there is no break or any signal of difference between habituation and test trials, we did not counterbalance the order of test trials – violation trials were always presented first. This raises the potential concern that infants may dishabituate to the first test trial, regardless of its content. Critically, however, this cannot explain the different patterns of dishabituation between the Approach and Avoidance condition, and specifically, the lack of dishabituation to the first violation test trial in the Avoidance condition.

Furthermore, to verify that infants did not dishabituate on the first violation test trial only because it came first, we conducted pairwise comparisons, treating the last habituation trial each infant received as if it were the first expected test trial. Thus, the “1st pair” analysis in this case compares the last habituation trial and the first violation

	Approach 7 month olds (Exp. 1a)	Avoidance 7 month olds (Exp. 1b)	Avoidance 14 month olds (Exp. 2)	Inherent Motion 7 month olds (Exp. 3)
Violation 1 – Avg. of Last 3 Hab (n = 16)	9.18 ** $t(15)=3.02$ $d=0.95$	-0.90 $t(15) = -0.48$ $d=0.12$	2.76 $t(15) = 1$ $d=0.26$	7.81 $t(15) = 1.56$ $d=0.39$
Violation 1 – Expected 1 (n = 16)	8.43 ** $t(15) = 3.78$ $d=0.95$	-5.73 † $t(15) = -1.96$ $d=0.49$	1.16 $t(15) = 0.32$ $d=0.08$	7.84 $t(15) = 1.58$ $d=0.40$
Violation 2 – Expected 2 (n = 15)	3.69 * $t(14) = 2.31$ $d=0.60$	-0.94 $t(14) = -0.92$ $d=0.24$	-2.95 $t(14) = -0.76$ $d=0.20$	10.71 † $t(14) = 1.90$ $d=0.49$
Avg. Violation – Avg. Expected (n = 16)	6.17 *** $t(15) = 4.32$ $d=1.08$	-3.45 * $t(15) = -2.49$ $d=0.62$	-0.34 $t(15) = -0.13$ $d=0.03$	9.45 ** $t(15) = 3.52$ $d=0.88$

Table 2.1: Differences between means of the average looking times on the trials indicated. Significance was obtained from planned paired-sample t-test comparisons, 2-tailed. († $p < .1$; * $p \leq .05$; ** $p < .01$, *** $p < .001$). d-values are point estimates of Cohen's d measure of effect size.

In the Approach condition infants looked significantly longer at violation trials than expected trials, both for the first pair of test trials and for the second pair, indicating that they expected the hand to continue reaching for the fixed object. This conclusion is further supported by their longer looking on the first violation test trial than on to the last three habituation trials. However, in the Avoidance condition infants did not look longer at violation trials than expected trials; in fact they trended in the opposite direction. Nor did they recover interest to the first violation test trial relative to the last three habituation trials. Thus, infants in the Avoidance

test trial. The 2nd pair comparison is between the first expected and the second violation test trial. And the comparison over the averages of violation and expected trials would take the average of the first and second members of these pairs. Analyzing the data in this way reveals essentially the same pattern of findings. Importantly infants in the Approach condition looked longer at the violation trials, $t(15)=3.15$, $p=0.007$, averaged across both pairs, whereas infants in the Avoidance condition did not differentiate, overall, between the two trial types, $t(15)=1.51$, $p=0.15$. In the Avoidance condition, infants looked significantly longer at the expected trial that follows the first violation trial relative to that violation trial, $t(14)=2.65$, $p=0.02$, an effect discussed in the text.

condition apparently failed to expect that the hand would continue not reaching for the fixed object. Nonparametric analyses confirmed these results. Fourteen out of 16 infants looked longer at the violation than the expected test trials in the Approach condition, which was significantly different from chance according to a Wilcoxon Signed Ranks Test ($z = 2.99$, $N = 16$, $p < 0.01$). In contrast, 4 out of 16 infants looked longer at the violation events in the Avoidance condition, which was also different from chance, but in the other direction ($z = -2.17$; $N = 16$, $p = 0.03$); that is, infants looked longer at the expected outcome.

2.2.4. Discussion

In the Approach condition infants expected a hand that consistently reached for one object to continue reaching for that object. In contrast to the basic Woodward paradigm, the setup of the Approach condition rules out the possibility that infants' dishabituation depended upon encoding a consistent avoidance of one of the objects. Rather, consistent with previous interpretations, they apparently took the repeated approach to the fixed object as evidence for a positive disposition toward it, and encoded the fixed object as the hand's goal, showing surprise when the hand approached, for the first time, one of the variable objects.

In contrast, infants in the Avoidance condition did not form an expectation that the hand's goal was always the variable object, and never the fixed object, or did not attribute a negative disposition toward this object to the agent. In fact, in the Avoidance condition, infants looked longer at the first expected than the first violation trial. That is, they looked longer at the trials when the hand reached for a novel variable object, even though they had been habituated to the hand always reaching for the variable object rather than the fixed object (e.g., a ball). This counterintuitive result has a number of possible explanations. It may reflect a powerful tendency to impute approach goals to agents. Recall that the first expected trial always followed the first

violation trial. Infants may be rapidly encoding the fixed object as a goal on the first test trial, when the hand reaches for the fixed object (e.g., the ball) for the first time. This reach towards the ball does not surprise infants, consistent with our interpretation that they had failed to encode the habituation trials as the agent avoiding the ball. Next, on the second test trial the hand reaches for some entirely new object, now avoiding the ball. For the first time in the experiment, infants are confronted with a non-reach to a previously reached-for object – the ball – that is still onstage. This may explain greater interest on the second and fourth test trials – the trials that would be “expected”, if infants represented the habituation events in the Avoidance condition as consistent avoidances, which apparently they did not. Alternatively, infants may have represented the identity of the object that is reached for on every trial, without forming a representation of the agent’s goal from a single trial. They may have learned a pattern : after an object is reached for, it is replaced. The first expected trial presents a violation of this pattern, which causes infants to dishabituate.

2.3. Experiment 2

Although 7-month-olds already have sophisticated representations of others’ intentional states, these representations continue to be greatly enriched throughout development. Between 9 and 15 months, infants increasingly show a capacity for joint attention and attention-following, among other skills (Carpenter, Nagel, & Tomasello, 1998). By 10 months, but not at 9, infants expect agents to look at another agent they are interacting with (Beier & Spelke, 2012). Representations of others’ goals in particular are enriched as well. By 12 months, but not at 7, infants interpret the direction of an agent’s gaze (Woodward, 2003) and pointing behavior (Woodward & Guajardo, 2002) as object-directed. Fourteen-month-olds, but not 10-month-olds, look predictively in anticipation of the outcomes of goal-directed reaching events in which

objects were moved around, but not to similar hand movements with a closed fist (Gredebäck, Stasiewicz, Falck-Ytter, Rosander, & van Hofsten, 2009). Many other studies have not explicitly compared younger and older infants, but have reported abilities in 12-14 month-olds that have not been reported in infants any younger. Infants at these ages differentiate between an agent who is unwilling versus one who is unable to complete a goal-directed giving action (Behne, Carpenter, Call, & Tomasello, 2005) and, knowing two out of the three components of the teleological analysis of rational actions (environmental constraints, goals, means), infer the third (Gergely & Csibra, 2003).

Unlike older infants, 7-month-olds also have severely limited domain-general information processing capacities. For example, working memory increases greatly between 6.5 and 12-months, probably due to maturation of prefrontal cortex (Diamond & Goldman-Rakic, 1989). Other executive functions, such as the capacity to inhibit prepotent responses also show marked developmental changes over this age range (Diamond, 1991). Relatively impoverished goal representations, plus domain-general information processing limitations, may have contributed to the failure of the 7-month-olds in the avoidance condition. Perhaps their failure reflects a relatively superficial bias in favor of approach over avoidance representations as well as information processing limitations that preclude overcoming this bias. If this is so, much older babies, relative experts at goal representations, with markedly increased executive function, might succeed at the avoidance condition of Experiment 1. Experiment 2 tests this hypothesis with 14-month-olds.

2.3.1. Method

2.3.1.1. Participants

Sixteen full-term infants (mean age 14;1, range 13;18-14;22, 5 female) participated in Experiment 2. Three additional infants were tested, but were excluded from analysis due to fussiness.

2.3.1.2. Procedure

The stimuli, setup and procedure were identical to the Avoidance condition of Experiment 1.

2.3.2. Results

As in Experiment 1 we excluded data when the difference between the violation and expected test trials exceeded 2 standard deviations from the mean difference across subjects, considering the first and second pairs of test trials separately. Two pairs of second test trials were excluded on this basis. The results are shown on Figure 2.2.

2.3.2.1. Habituation

Nine infants reached the habituation criterion, and a paired samples t-test indicated that infants looked less over time during habituation (M first 3 = 32.2; M last 3 = 15.8; $t(15) = 3.96$, $p = 0.001$). Comparing the 14 month old infants in Experiment 2 to the 7 month olds in the Avoidance condition of Experiment 1, an ANOVA examined the effects of experimental group (7 m.o. vs. 14 m.o.) and habituation trial (first 3 vs. last 3) on looking times during habituation. There was no significant interaction, $F(1,30) = 0.87$; $p = 0.36$, suggesting that infants of different ages did not habituate at different rates to the Avoidance display; There was a significant main effect of experiment, $F(1,30) = 8.23$; $p < .01$, indicating that the older infants looked longer over the habituation phase (M 14 m.o. = 24.01; M 7 m.o. = 14.7).

2.3.2.2. Test Trials

An ANOVA examined the effects of the within-subject variables of trial type (Violation vs. Expected), trial pair (first vs. second pair) and experiment (7 m.o. vs. 14 m.o.) on looking times during the test trials. This revealed a main effect of trial pair, $F(1,28) = 4.5$, $p = 0.04$, with infants looking longer overall on the first pair of test trials. There was also a main effect of experiment, $F(1,28) = 9.8$, $p < 0.01$, with older infants looking longer than younger ones at the test events. Importantly, there was no significant interaction of trial type and experiment, suggesting that older infants did not differentiate the expected from the unexpected events any more than did younger infants. Planned comparisons (see Table 2.1) found that the 14-month-olds failed to dishabituate to the first violation test trial, relative to the last three habituation trials, and did not differentiate the violation test trials from the expected test trials, nor on the first or second pairs of test trials, nor overall.¹³ Nine out of 16 infants looked longer at the violation test trials than the expected trials, which was not significantly different from chance according to a Wilcoxon Signed Ranks test. Thus, 14-month-olds, like 7-month-olds, failed to predict that the hand would continue to avoid the fixed object, in favor of the variable ones.

2.3.3. Discussion

Like younger infants, 14 month-olds do not look longer at violation than expected events in this Avoidance condition. Given all the additional capacities of 14-month-olds over 7-month-olds, both domain general capacities such as increased working memory and executive function (Diamond, 1991), and domain specific capacities relevant to intentional attributions (Gredebäck, et al., 2009; Beier & Carey, 2013; Woodward, 2003, Carpenter, et al., 1998, Behne, et al., 2005;

¹³ As in Experiment 1, we again analyze the data in a way that treats the last habituation trial as if it was the first test trial. So the “1st pair” analysis here compares the last habituation trial and the first violation test trial. The 2nd pair comparison is between the first expected and the second violation test trial. And the comparison over the averages of violation and expected trials would take the average of the first and second members of these pairs. Analyzing the data of Experiment 2 this way, there were no significant differences between violation and expected trials in Experiment 2.

Gergely & Csibra, 2003), the older infants' failure to represent avoidances in the present study is particularly striking.

We do not claim that 14-month-old infants cannot represent avoidance behavior under any circumstances, since it is of course possible that they would succeed if given additional cues that the agent's goal is in fact to avoid the fixed object. The same is true of 7-month-olds. Further work is required to characterize the extent of infants' difficulty with avoidance and the conditions necessary for them to represent avoidance goals. For example, if the agent showed a negative emotional reaction towards the avoided fixed object, or if the hand hovered over the avoided object repeatedly, as if considering reaching for it, before reaching for the variable object every time, infants may be more likely to notice the avoidance. Nevertheless, the experiments presented here show that there is a significant asymmetry between the evidence infants need to represent a goal-directed approach versus an avoidance. While the evidence given 7-month-olds in the Approach condition suffices to lead them to consistently expect continued approach, perhaps having taken consistent approach as evidence for a positive disposition toward the fixed object, the same sort and amount of evidence is not enough for 7-month-olds, or even 14-month-old infants to expect continued avoidance.

2.4. Experiment 3

Experiments 1 and 2 show that infants do not encode consistent reaches to and away from an object with equal ease, given the same kind of evidence. It is possible that on the basis of observed actions, avoidance goals are more difficult to encode than approach goals because of the basic structure of infants' capacity for goal inference. We will return to why that might be so. However, infants may also have failed to encode the consistent avoidance in experiments 1 and 2 because of general cognitive demands of the Avoidance condition that are not specific to

representations of the agent's attitudes toward the objects or to the agent's goals. For instance the movement of each object may be an overpowering attentional draw, such that on every trial infants encode only the moved object and fail to encode the object that remains still on the stage. This would lead to encoding the identity of the fixed object in the Approach condition, since it is the one that always moves, and failure to encode the identity of the fixed object in the Avoidance condition, since it never moves. Relatedly, it may be that the cognitive resources necessary to represent what is happening to the bear preclude an additional representation of what *isn't* happening to the ball. Or perhaps infants represent the entire Avoidance display as, "a different object moves every time" – a representation not possible in the Approach condition – in which case the violation test event is not a violation of the rule and would not be expected to cause dishabituation. Finally, perhaps infants simply assume that the fixed object is part of the plank, and thus don't attend to it at all.

Experiment 3 tests these alternative accounts. Instead of a hand reaching for each variable object, we draw the infant's attention to the variable objects by making the objects move on their own, rising into the air and dangling (pulled up on invisible strings controlled by the experimenter). The fixed object never moves during habituation, moving only on the violation event during the test phase. If the attentional pull or processing demands of the moving object preclude attention to or inferences about the stationary object, then 7-month-old infants will fail to dishabituate to the violation event here as well. In contrast, if the failure in Experiments 1 and 2 reflects constraints on representing intentional actions per se, 7-month-olds may succeed in Experiment 3, where the inherent motion of the objects does not signal any goal-directed action. Since there are no reaching actions, there is no avoidance to represent in this case, and no dispositional attitudes of an agent towards the object. Infants would need to encode only that

there is one exceptional object, which sits still while all the others move, in order to dishabituate when that fixed object finally moves on the violation test event.

The stage and object setup of Experiment 3 was identical to the Avoidance condition of Experiment 1. Variable and fixed objects were attached to thin wire which allowed them to be lifted and dangled in the air (exhibiting motion similar to Experiment 1, but without a hand). Following the design of the Avoidance condition of Experiment 1, the variable object dangled in the air on every habituation trial, while the fixed object never did. Thus, Experiment 3 was identical to the Avoidance condition of Experiment 1 except that the motion of the objects was not caused by a visible hand.

2.4.1. Method

2.4.1.1. Participants

Sixteen full-term infants participated in the study (mean age 7;3, range 6;15-7;15, 7 female). Ten additional infants were excluded and replaced: 9 due to fussiness and 1 due to parental interference.

2.4.1.2. Stimuli and Setup

We modified the stimuli used in Experiment 1 by gluing a small magnet on top of each object. While objects were onstage, this magnet was connected to a thin wire, allowing the experimenter to lift and dangle each object by tugging on the wire from above (out of the infant's view). Once each variable object was shuttled offstage, the experimenter disconnected it from the wire and swapped a new variable object into place.

2.4.1.3. Procedure

The procedure was identical to Experiment 1, except that instead of a hand reaching for an object in the infant's view, that object was lifted by the wire and dangled, jiggling, until the

infant looked way and the trial ended. We attempted to give the strongest possible impression of a freely floating object by using black wire against a black background.

2.4.2. Results

2.4.2.1. Habituation

The results are depicted on Figure 2.1 and in Table 2.1. Looking times decreased significantly during habituation (first 3 trials $M = 28.8$; last 3 trials $M = 15.09$; $t = 2.55$, $p < .05$). Nine infants in Experiment 3 reached criterion, which was significantly fewer than the 14 infants who reached criterion on the Avoidance condition of Experiment 1 ($\chi^2(1, N = 32) = 3.87$, $p < 0.05$). However, an ANOVA comparing habituation (first vs. last three trials) between this experiment and the Avoidance condition of Experiment 1 (Inherent Motion vs. Avoidance) found no interaction, $F(1,30) = 0.07$, $p = 0.79$. Infants, therefore, entered the Test Trials of Experiment 3 having lost interest in the habituation trials to a similar degree as infants in the Avoidance condition of Experiment 1. There was also a marginal main effect of condition, $F(1,30) = 3.76$, $p = 0.06$. Infants' overall looking to the free-floating objects was greater than to those lifted by hand, perhaps because they do not expect objects such as these to move by themselves and float.

2.4.2.2. Test Trials

One infant's second pair of test trials was excluded from analysis because the difference in looking time between expected and violation trials was greater than two standard deviations from the mean. As can be seen from Table 2.1, averaging across the first and second pair of test trials, infants looked significantly longer to violation trials than expected trials $t(15) = 3.52$, $p < 0.01$.¹⁴ This pattern was observed on both the first and second pairs of test trials, although not

¹⁴ As in Experiments 1 and 2, we again analyze the data in a way that treats the last habituation trial as if it was the first test trial. So the "1st pair" analysis here compares the last habituation trial and the first violation test trial. The

significantly on each pair. A non-parametric analysis confirmed this result. Fourteen out of 16 infants looked longer at the two violation trials than the two expected events, significantly more than expected by chance according to a Wilcoxon Signed Ranks test ($z = 2.69$, $N = 16$ $p < 0.01$). Thus, infants were more surprised to see floating by a previously non-floating object, compared to floating by a novel object.

To confirm that the difference in the pattern of results from the Avoidance condition in Experiment 1 and those from Experiment 3 was statistically reliable, an ANOVA examined the effects of trial type (expected vs. violation), trial order (first pair vs. second pair), and Experiment (Avoidance condition of Experiment 1 vs. Experiment 3) on looking times during the test trials. This analysis revealed a significant interaction of trial type and condition, $F(1, 29) = 14.45$, $p = .001$, and no other significant main effects or interactions. The interaction reflected the fact that infants looked longer at the unexpected test trials only in Experiment 3, the reverse pattern compared with the Avoidance condition of Experiment 1 (see Figure 2.2).

2.4.3. Discussion

In Experiment 3 infants looked reliably longer when an object rose and jiggled if that particular object had repeatedly failed to do so on previous trials, compared to when a novel object moved. Thus, infants were able to notice that the fixed object was always still and expected it to continue being still. This finding constrains our interpretation of Experiment 1, in which infants were not surprised to see a hand lift and jiggle an object when, previously, that object had never moved.

^{2nd} pair comparison is between the first expected and the second violation test trial. And the comparison over the averages of violation and expected trials would take the average of the first and second members of these pairs. Analyzing the data of Experiment 3 this way, there is a significant difference between the averages of the two expected trials, on the one hand, and the two unexpected trials, on the other: $t(15)=2.62$, $p=0.02$.

Many experiments show that even young infants may take the capacity for self-generated motion as evidence that an entity is a dispositional causal agent, whereas evidence that an entity moves only upon contact leads to its categorization as dispositionally inert (e.g., Luo & Baillargeon, 2005; Saxe, Tenenbaum, & Carey, 2005; Saxe, Tzelnic, & Carey, 2007; Muentener & Carey, 2010; Luo, Kaufman & Baillargeon, 2009). That is, both of the categories dispositional causal agent and inert object are available to infants, and both have inferential consequences in predictions about future events that involve that type of entity. Thus, it seems possible that the intermittent self-generated movement of each variable object provided a sufficient cue to it falling in the former category, while the still object was more saliently inert by contrast. Infants' surprise at the violation trial of this experiment may be due to an object that they thought was inert moving on its own for the first time, effectively revealing itself to be a causal agent. Nevertheless, in order to make such an inference, infants must have attended to the fixed object and represented its inertness. Thus, infants' failure in the Avoidance condition of Experiments 1 and 2 cannot reflect a complete inability to process information about persistent inertness, given infants' success in Experiment 3. Nor can it reflect an inability to attend to the non-moving fixed object during habituation, nor their representing these situations as "a different object moves every time", nor discounting the fixed object as affixed to the plank and thereby irrelevant, nor an inability to entertain two simultaneous representations – both of what is dangling and what is, by contrast, still. All these would have predicted a failure on Experiment 3. Rather, infants' failure in the Avoidance condition of Experiment 1 and Experiment 2 may reflect a more narrow processing bias or constraint, specific to the domain of interpreting goal-directed behavior and agent's positive or negative dispositions toward objects.

It is possible that a hand's picking up an object and shaking it is a more powerful attentional draw than is that object's moving by itself, such that the failure to dishabituate in the avoidance condition does reflect a failure to attend to the stationary object after all, in spite of infants' attending to this object in Experiment 3. Several considerations speak against this explanation. Previous research shows that the standard reaching paradigm only works when the hand reaches in a goal directed manner (not in a "flop"), suggesting that attentional biases towards hands or reaching things generally cannot account for infants' successes (Woodward, 1999). Additionally, infants had ample opportunity to attend to the avoided object in the Avoidance condition: it was present on-stage for the full duration of the experiment (including between trials, when no hands were present), it was never occluded from view, and it was shuttled across the stage between each trial. Finally, overall looking was greater in Experiment 3 than in the Avoidance condition of Experiment 1, suggesting that various objects moving on their own are a powerful attentional draw. We do not deny, of course, that the reaching and grasping hand in Experiments 1 and 2 is the relevant difference between these experiments and Experiment 3, for this is what sets up the context for representations of approach and avoidance, and positive or negative dispositions towards objects.

2.5. General Discussion

Experiment 1 demonstrates that 7-month-old infants expect consistency in the approach behavior of a goal-directed hand, but not in a hand's avoidance behavior. If an object has been consistently approached in favor of diverse others, infants expect the hand to continue to approach that object. In contrast, if an object has been consistently avoided in favor of diverse others infants do not predict that the agent will continue to avoid that object. Experiment 2 shows that this failure to generalize across instances of avoidance extends at least until 14 months.

Experiment 3 shows that if the objects move on their own instead of being picked up by a hand, 7-month-olds do expect an object that has never moved to continue being still, thereby controlling for a number of low level alternative explanations. These findings allow us to specify more precisely the contents of infants' early representations of others' mental states, particularly in the most commonly used paradigm to study early goal attributions and attributions of agents' dispositions towards objects.

Our results also bear on a debate concerning infants' goal representations that derives from Luo and Baillargeon's (2005) finding that, if there is only one object present during habituation, infants do not expect the agent to continue approaching that object during a subsequent test event in which there is a choice between it and a second object. Luo and Baillargeon argued that infants' expectation of a consistent reach is contingent on seeing evidence of the agent's relative preferences during habituation – for example, that the agent specifically prefers the ball to the bear, not just that the agent wants the ball. Further evidence for this interpretation derives from the fact that the nature of the foil matters (it must be a different basic-level kind from the target of the reach; Spaepen and Spelke, 2007) and the fact that 7- and 12-month-old infants are sensitive to whether the foil is visible to the reaching agent (Luo and Johnson, 2009; Luo and Baillargeon, 2007). Luo and Baillargeon concluded that infants are representing a preference between two particular objects.

However, infants' success on the Approach condition of Experiment 1 supports Baillargeon's and her colleagues' revision of this hypothesis (Baillargeon, et al., 2014). In the violation test trials of the Approach condition, infants expected an agent to continue reaching for a ball over a brush (for instance), even though the brush had never appeared before and there was therefore no evidence of the agent's preference between these two specific objects. This shows

that infants do not require evidence of the agent's preference of one out of two specific objects in order to form an expectation about the agent's future reaches towards either object. If infants do represent the agent's preference between the two objects, it is a generalized preference of the form the agent prefers X to other things rather than the agent prefers X to Y. Alternatively and perhaps most simply, as Baillargeon et al (2014) now argue, infants can infer that the agent has a general positive disposition towards the object – a representation that the agent likes X, though in some contexts an alternative choice or another cue to the agent's disposition is needed to support this inference. Other recent results converge on this interpretation. Robson and Kuhlmeier (2013) found that infants who are habituated to a reach for X over Y – but not to a reach for X when presented alone – will dishabituate to a reach for a novel object Z over X, again suggesting that their expectation of continued reaching for X does not depend solely upon a representation of a preference of X over Y.

The news from the present results is the failure of both 7-month-old and 14-month-old infants to distinguish between violation and expected events in the test trials of the avoidance displays. This failure does not show that infants of these ages are incapable of encoding avoidances, or, alternatively, of attributing negative dispositions toward objects. Whether or not there exist other mechanisms that process avoidance, it is at least the case that avoidance is not spontaneously computed by the specific mechanism responsible for attributing goals or agents' dispositions toward objects given consistent reaches to one object over another. This is the case even when the amount of information available to compute avoidance is strictly equated with the amount of information available to compute approach.

There are two possible biases, which are not mutually exclusive, that might explain infants' failure. The first is a bias operating on representational inputs: This mechanism infers

goals based on the representational input “approaches object”, as compared with the input “avoids object”. Although by presenting exactly two objects on each trial our study was designed to equate the statistics relevant to computations of approach and avoidance, these constraints rarely apply in natural settings. Typically, a hand reaching for a ball clearly means to grab the ball, while a hand reaching away from a bear (and for the ball) might not have any particular designs concerning the bear at all; it might simply want the ball. More generally, the number of actions an agent is performing at any given time is dwarfed by the number of actions that agent is not performing, and trying to compute for each non-action whether it is goal-directed is neither computationally plausible nor a good way of figuring out what the agent is likely to do next. Thus, it is plausible that there is a bias to attend selectively to approaches when encoding goals. The second possibility is a bias operating on representational outputs: This mechanism preferentially infers mental states of the type “wants or likes object”, “has the goal to touch the object” as compared with the types “*doesn’t want or doesn’t like object*” or “has the goal to avoid the object”.

In either case, the question arises how adults are able to represent avoidance goals. In order to verify that adults possess this competency and will spontaneously deploy it in our experimental design, we randomly assigned 16 adults to watch videos of either the Approach or Avoidance conditions, where the length and number of habituation trials matched the means from the respective conditions in the infant-directed paradigm in Experiment 1. Pausing the video at the end of the habituation trials, we asked adults to say what is happening in the video and to predict what would happen next. Although answers on the first question were generally under-informative (e.g. “A hand reaches out and grabs one of the objects. The objects get replaced”), in answer to the second question, every participant in both conditions described the

hand's pattern correctly and predicted that the hand would continue its pattern once the video resumed. Following these two prompts, adults saw a violation trial and were asked if this trial was what they expected. Again, every participant but one expressed surprise at the violation, saying that, for example, the hand was "going nuts", "breaking with tradition", or "undermining" its pattern in a way that was "shocking". The one dissenting participant, who watched the Approach condition, had previously predicted that the hand would continue grabbing the same object but said he had "hoped" it would grab the other one next, "just to change things up". This participant said he was not surprised when the hand violated its pattern, because he was "hoping that would happen". Out of the eight adults who saw the Avoidance habituation display, six used some form of verbal negation to justify their prediction of what would happen next (e.g. the hand grabbed the new toy "and not" the old one, "never" touched the old object, "hadn't" touched it, or "wasn't" touching it). The remaining two described the hand as "avoiding" the object.

Perhaps the most significant resource available to adults is a facility with a combinatorial system of thought. This system can access the logical concept NOT, and can represent the avoidance condition as a series of events of the same kind – reaches for NOT the ball. However, the conceptual basis of adult's competence in reasoning about avoidance remains an important topic for future research, as does the developmental process by which this competence is acquired. As a next step, it will be important for future work to determine the age at which infants or toddlers begin to succeed in representing avoidance from the same kind of information they use to represent approach.

The failure of infants to note the violation in the Avoidance display shows not only that the sequence of consistent non-reaches in the Avoidance condition doesn't spontaneously trigger an avoid representation, but also that the approach representations that they do generate do not

spontaneously enter into a domain-general computation with a negation operator. (“She wants the ball but not the bear... She wants the hat but not the bear... She wants the apple but not the bear....”, etc.) This account predicts that success on the Avoidance condition of Experiment 1 would be enabled as children gain facility with truth-functional negation – the knowledge that if some proposition P is true, then its negation, NOT-P is false. Little is known about the developmental course of this capacity, but it does not appear in children’s speech until at least 24 months or, according to some, much later than that (24 months, Pea, 1982; around 28 months, Hummer, Wimmer and Antes, 1993; not until 4-5 years, Kim, 1985; Nordmeyer and Frank, 2014). While the findings of infants’ success in Experiment 3 might be seen as challenging to this hypothesis, they are far from showing that infants succeed in processing negation or attending to consistent non-events broadly. Rather, as we argued above, infants have categories of both self-moving dispositional causal agents and inert objects. If infants interpreted the variable floating objects as agents, they may have succeeded on Experiment 3 by noticing the contrast between these agents and the inert fixed object. Being surprised that what they thought was inert suddenly turned out to be an agent does not require use of a negation operator.

The distinction between approach and avoidance is related to the distinction between actions and omissions. The avoidances we presented to infants are a form of omission: The agent is not acting on the fixed object. The distinction between actions and omissions has been explored in the literature on adults, where it has been shown to influence adults’ judgments of causal responsibility, intent, and blame (Spranca, Minsk, & Baron, 1991; Ritov & Baron, 2010). For example, if adults read about Evan who deliberately tips over a bucket of water and Jeff who deliberately allows a shaky bucket to tip over on its own, they say that Evan caused and intended the water to spill more than Jeff did. (Cushman & Young, 2011). If they read about Evan who

deliberately kills a man and Jeff who deliberately allows a man to die, they say that Evan committed a greater moral wrong (e.g., Spranca, et al., 1991).

Our findings suggest a possible connection between infant cognition and the adult bias to draw stronger inferences from representations of actions than from representations of omissions. Adults appear to place observed actions into a causal and moral framework spontaneously, but require some measure of cognitive control to treat omissions the same way (Cushman, Murray, Gordon-McKeon, Wharton & Greene, 2011). This suggests continuity across infants' and adults' representations of events. If infants' capacity for goal attribution operates over actions more readily than over omissions – or if, in the extreme case, the capacity to engage in the controlled processing needed to see omissions and avoidances as goal-directed emerges later in development – the adult “omission effect” might reflect an enduring signature of the earliest-emerging psychological capacity for intentional attribution.

Chapter 3:

What do you mean, “No”? Young children’s production of some negation words precedes their understanding of logical negation

3.1. Introduction

One of the most remarkable things about human cognition is the ability to take a finite number of concepts and combine them to generate an infinite number of meaningful thoughts. An English speaker who has never heard the sentence or entertained the thought, THERE ARE NO BEARS ON MARS, has no trouble understanding what it means. Not only can we understand it, we are able to judge that it is very likely true and make conclusions on that basis: if there are no bears on Mars, that means there are no brown bears there, no bear cubs, no bears climbing Martian trees. The ease with which a language user can understand new thoughts, make judgments about their truth, and reason through to related thoughts, all have just one plausible explanation – the thinking of complex thoughts is the result of a rule-governed system that combines meaningful component units (words, concepts) in systematic ways.

To understand the meaning of a sentence entails understanding its truth-conditions – that is, under what circumstances it would be true or false (Heim & Kratzer, 1998). And to understand the meaning of a sentence containing a negative like, THERE ARE NO BEARS ON MARS entails understanding that its truth conditions are the opposite of THERE ARE BEARS ON MARS. Negation, as expressed in English by the words “No” and “Not”, is an operation that flips the truth value of a sentence – if an affirmative sentence is false, that sentence negated is true, and vice versa. This makes negation a highly abstract function word. Unlike the vast majority of content words – nouns, verbs and adjectives – it does not refer to anything in the world, or even in the mind (like BELIEF or WANT). Its only meaning is in the systematic way in which it alters the meanings of the propositions it combines with. Yet despite how abstract the negation operator is, there is not a language in the world without a word for it and not a culture that has ever been reported to dispense with it in their thought.

Taking negation as a case study of abstract, combinatorial concepts, this paper explores its developmental origins. We ask whether and how learning the language of negation might be involved in acquiring the concept of this truth-functional logical operator – that is, a mental symbol with the conceptual role of the negation operator and the right combinatorial properties to meaningfully compose with propositions and other concepts. Do children have a concept of truth-functional negation before they learn the words “No” and “Not”, so that learning these words is a matter of mapping a label to a preexisting concept? Or does learning the word precede learning the concept? If so, is it because children do not have the concept at the time they first learn the word, or because mapping the concept to the word is a difficult language-learning problem? Does the word perhaps participate in acquiring the concept in some way, for example by creating a linguistic placeholder structure upon which increasingly rich and abstract representations can later be built (Carey, 2009)?

To make progress on these questions, we look at two aspects of the linguistic development of negation, comparing the developmental courses of the production of the words “No” and “Not” to the comprehension of the logical meaning of those words. The idea is to look for a gap between the two. If we find that for each word logical comprehension and production track together, it would suggest that the child has the concept before the word and that mapping the word to the concept is constrained primarily by the input frequency of each word and by their different grammatical properties. In this case, as with many other words, comprehension might also precede production as children gradually gain certainty that they have made the right mapping between word and concept (Goldin-Meadow, Seligman, & Gelman, 1976; Harris, Yeeles, Chasin, & Oakley, 1995).

The third possibility is that the production of negation words might precede their comprehension as logical operators. Although less common than comprehension preceding production, this pattern of acquisition has been observed in word learning across a number of domains, including color words (Pitchford & Mullen, 2003; Sandhofer & Smith, 1999; Soja, 1994; Wagner, Dobkins & Barner, 2015), object labels (Ameel, Malt & Storm, 2008), number words (Wynn, 1992; Carey, 2009), time words (Tillman & Barner, 2015; Busby Grant & Sudendorf, 2011; Shatz, Tare, Nguyen, & Young, 2010), certain theoretical terms, like ALIVE and DEAD (Carey, 1985), and emotion words (Widen & Russell, 2003). Whenever production of a word precedes comprehension of the meaning it has for adults, there are two classes of explanations possible. One is that the child does not yet have the adult-like concept at the time when the word is first learned. This is the case in which learning the word could possibly play some causal role in the acquisition of the concept. The other possibility is that the adult concept is present, but that mapping the word to that concept is itself a hard problem because the hypothesis space of possible mappings is initially insufficiently constrained for the child to identify the right one (see Snedeker and Gleitman, 2004 for discussion).

Consistent with the possibility that production of negation words precedes comprehension of the truth-functional meaning of negation, longitudinal studies have found that when children first begin to produce these words, they appear to use them for a narrower set of functions than do adults (Bloom, 1970; Pea, 1980a). Despite the fact that both “No” and “Not” are equally common in their input (at least when a reliable quantitative estimate is available, at 25 months; see Cameron-Faulkner, Lieven & Theakston, 2007), children begin to produce “No” much earlier than “Not” (Dale & Fenson, 1996; see Fig. 3.1), and the types of meanings they express with “No” are initially more limited than the meanings used by adults. At first, when

children are producing mostly single-word utterances, they use single-word “No” primarily to reject prohibitions, imperatives and other forms of parental tyranny. When they begin to combine “No” with other words to form two-word utterances, the primary use of negation shifts to express ‘nonexistence’ – the often unexpected absence of the other word’s referent, as when a child opens the fridge to look for juice, finds none and says, “No juice!” Although estimates vary depending on whether one looks at the earliest age, the age of consistent or frequent use, or some point in between, it is not until 24-30 months that children systematically use “No” to deny the truth of others’ utterances (Bloom, 1970; Pea, 1980b, 1982; Hummer, Wimmer and Antes, 1993; Choi, 1988). These are cases when a child points to a dog and says, “Not bear”, or when a parent asks, “Is this a bear?” while pointing to a dog and the child answers, “No”. This latest-emerging function is called ‘denial’, and is most closely associated with truth-functional negation (Pea, 1982). When children first begin to produce denials, they use both “No” – which had previously been used to express nonexistence and rejection – and also begin to use “Not”. The emergence of “Not” in a child’s productive vocabulary tracks closely with the emerging production of denial negations with either word, although once it begins to be produced, “Not” is also used to express other functions (Choi, 1988; Cameron-Faulkner, et al., 2007). Given that children begin to produce denial negations between 24 and 30 months, why not simply conclude that this is the age at which children map the concept of truth-functional negation to both the words “No” and “Not”?

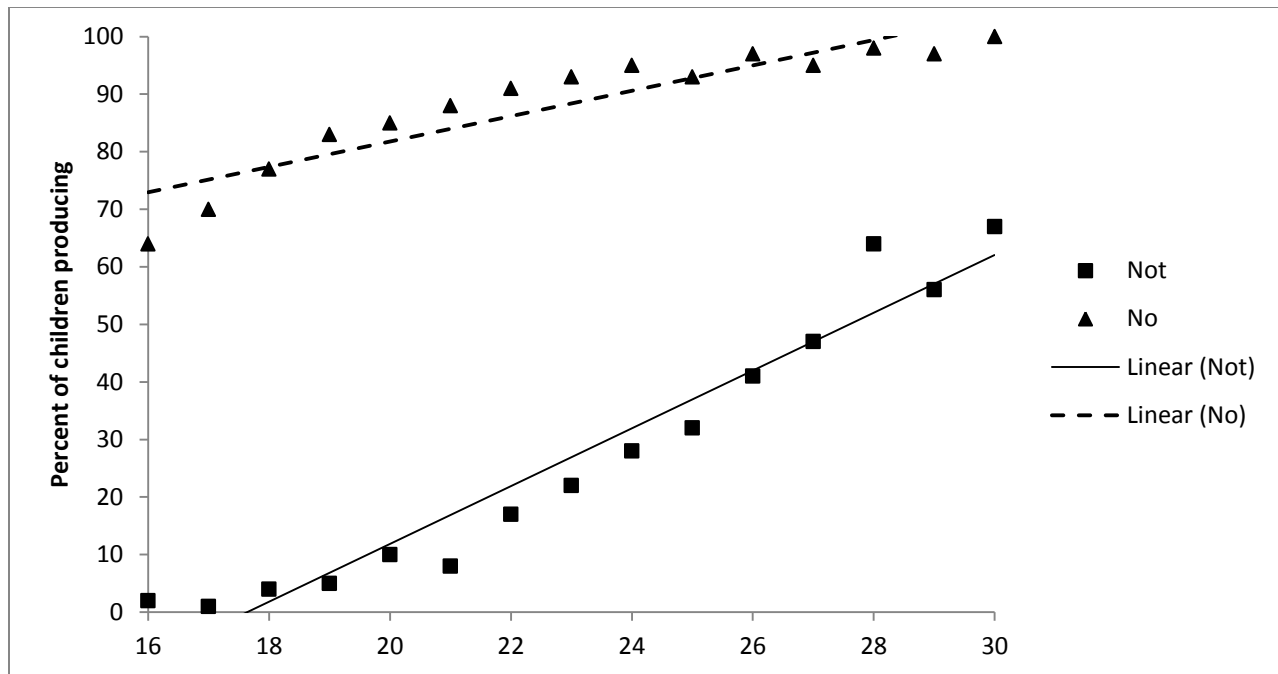


Figure 3.1: Production of the words “Not” and “No” by age, from the Wordbank database of Macarthur-Bates CDI norms. Children’s age in months is along the X-axis, and the percent of children producing the words is on the Y-axis.

The problem with data from production alone is that it is impossible to distinguish different meanings from different uses. Adults, who understand the truth-functional meaning of “No”, can nevertheless use it to reject or signal nonexistence as well as to deny. It is easy enough to paraphrase the narrower meanings of ‘rejection’ and ‘nonexistence’ as fully composed propositions containing truth-functional negation. A rejection “No!” means I DO NOT WANT THIS and a nonexistence “No juice” means THERE IS NO JUICE IN THE FRIDGE. Perhaps children’s relatively narrow uses of negation reflect the kinds of speech acts they prefer to make rather than the kinds of meanings they have mapped to negation words. Perhaps a young child is in a position where she knows little about the world, but is confident about her distaste for bathing, and therefore has less occasion to deny a factual claim than to reject the offer of bath time, even if she does understand negation to be a truth-functional operator. To address this ambiguity, we

turn to examine children's comprehension of negation. If, for young children, negation really does not have a truth-functional meaning, they would fail in comprehending sentences where negation words invert truth value, even though they might be producing those same words.

Though there are fewer studies of the comprehension of negation than of its production, those that exist point to the comprehension of truth-functional negation posing a significant challenge for children much older even than the age at which they begin to produce denial negations. Kim (1985) presented 3-5 year-old children with a puppet that would sometimes describe objects correctly and sometimes incorrectly, using both affirmative and negative statements. Given a banana, the puppet might say, "This is a banana", "This is not a banana", "This is an apple", or "This is not an apple". Using a truth-value judgment task, Kim asked children to judge if the puppet was right or wrong. Children from three to five years old correctly judged true affirmative sentences as 'right' and false affirmatives as 'wrong', but struggled with both false and true negatives. Children had the most difficulty with true negatives: even five-year-olds judged them as correct only 62% of the time (see also, Lloyd & Donaldson, 1976; Donaldson, 1972). Although it is possible to interpret these failures as reflecting a lack of a conceptual understanding of negation, there are several other reasons children might have difficulty with this task. The task is metalinguistic, with children having to judge whether a statement is right or wrong rather than responding directly on the basis of affirmative or negative information. Not only might this kind of judgment pose an added layer of difficulty, it also complicates the interpretation of children's responses. Judgments of whether a statement is 'right' or 'wrong' depend not only on truth value, but also on considerations of pragmatic felicity (Pea, 1982; Katsos & Bishop, 2011). Pragmatic felicity poses a particular problem for true negative sentences. Saying that a banana is not an apple when no apples had previously been

mentioned may be reasonably judged as conversationally inappropriate, even if children know that it is, strictly speaking, true. The paper you are reading is not an aardvark, but processing that proposition involves considering aardvarks and their possible relation to psychology research papers for the first time in this discourse. Using negated statements to introduce novel concepts into the discourse out of the blue imposes a processing load on adults (Kaup, Ludtke, & Zwaan, 2006; Kaup, Yaxley, Madden, Zwaan, & Ludtke, 2006) and may cause failure of ultimate comprehension in children, even outside of a metalinguistic truth value judgment task.

Recently, Nordmeyer and Frank (2014) took a less metalinguistic approach, looking at children's spontaneous eye movements as they were presented with two pictures – a boy with apples and a boy with nothing (or, in a different condition, a boy with gifts). They asked children between the ages of two and four to “look at the boy who has no apples”. All children started out by looking at the boy with apples. While older children eventually recovered to look at the right boy (at above-chance levels in one of the conditions, if still not perfectly) two-year-olds never looked away from the boy with apples. On the one hand, this too is consistent with an inability to understand truth-functional negation at age two. On the other, it is also consistent with an imperfect sentence processor that considers the predicate BOY WHO HAS APPLES, finds its referent, and gets stuck there. As Nordmeyer and Frank point out, it would be surprising if kids all through the third year of life truly failed to understand logical negation, given that these children produce ‘denial’ negations consistently (Hummer, et al., 1993; Pea, 1980a, 1982).

Finally, just one recent study of children's comprehension finds success at an age more consistent with their production of denial negations. Austin, Theakston, Lieven, & Tomasello (2014) get around the limitations of previous paradigms by presenting children with a pragmatically supportive context where both the negative and affirmative statements they hear

are made plausible by the preceding discourse. Children play a game in which they have to find a ball that an experimenter has hidden in either a bucket or a house. A second experimenter asks one of two questions – “Is it in the house?” or “Is it in the bucket?” – and the first experimenter answers differently across three different conditions. In a ‘single word’ condition, the experimenter says simply “Yes” or “No”, in a ‘sentence’ condition, the experimenter says, “It’s in the [bucket/house]” or “It’s not in the [bucket/house]”, and in a ‘gesture’ condition, the experimenter either nods to indicate an affirmative or shakes their head for a negative. Austin, et al. find earlier success with the comprehension of negation than previous studies, with 24-month-olds basing their search on negative verbal information, and 29-month-olds doing so even more reliably – that is, looking in the correct location when given either the single-word clue, “No”, or the sentence clue, “It’s not in the bucket”. This age of success tracks closely with children’s emerging production of the word “Not” (Bloom, 1970; Pea, 1980b, 1982; Choi, 1988; Hummer, et al., 1993), suggesting that the production of denial negations does reflect an emerging concept of truth-functional negation.

In the search for a gap between the production of negation words and their comprehension as logical operators, the findings of Austin, et al. (2014) provide a tentative answer. They find that children begin to comprehend “No” and “Not” as logical operators at the same time, and that this happens at the same age range at which children typically begin to produce “Not” and to use both “No” and “Not” to deny. This indicates a large production-comprehension gap for “No”, and no gap for “Not”. However, some aspects of Austin et al.’s data give pause to this conclusion. To negate a proposition, one must be able to represent the content of the un-negated proposition and combine it with the negation operator. The negative sentences used in Austin, et al. are the negated counterparts of their affirmatives. It is therefore

surprising that, while the children they tested performed significantly better than chance with negative clues containing both “No” and “Not” at 24 months, both 24- and 28-month-olds did no better than chance when given whole-sentence affirmative clues (e.g. “It’s in the bucket”). This pattern of results is even more surprising in the context of rich evidence of younger children’s robust understanding of affirmative sentences across a variety of other paradigms (Kim, 1985; Hummer, et al., 1993, Pea, 1980b; 1982).

That the findings of Austin, et al. underestimate children’s comprehension of affirmatives raises the possibility that they also underestimate their success with negatives. It is therefore important to provide converging evidence for the age at which children understand the negation words “No” and “Not”, ideally in a task where children respond successfully to affirmative statements as well. In Austin, et al.’s paradigm, affirmative and negative trials alternate. This may cause interference from one trial type to the other. To facilitate success on affirmative – and possibly also negative – trials, we use a blocked design. In case children’s ability to use either type of information is obscured by their perseveration from one type of trial to another, a blocked design may reveal a latent competence, even if only on the first block. We also go beyond identifying an average age of success in the comprehension of truth-functional negation and relating it to an average age of production that is based on population norms. By collecting productive vocabulary measures for each child we test, we can directly relate comprehension success to each child’s developing production of the words “No” and “Not”.

If children have the logical concept before they learn these words and mapping the words to the concept has to do with getting sufficient exposure to each word, there should be no gap between production and comprehension for either word and no relationship between the acquisition of the two words to each other. In this case, children should begin to comprehend the

truth-functional meaning of “No” much earlier than of “Not”, since they produce “No” nearly a year earlier (see Fig. 3.1). If children successfully comprehend logical “No” and “Not” at the same time – which, since “No” is produced earlier, would translate to a larger comprehension-production gap for “No” and a small or no gap for “Not” – it would suggest a single limiting factor for the logical comprehension of both “No” and “Not”. There are several possibilities for what that could be. It may be the concept of truth-functional negation, with acquisition of that concept enabling a relatively problem-free mapping for both “No” and “Not”. It could also be that mapping the words “No” and “Not” to a logical concept requires knowing enough language – knowing the syntactic roles of words and how they combine along with the syntactic properties of the negative particle, as well as knowing the meanings of a sufficient number of content words. When “No” or “Not” are heard in a sentence, the meaning of the sentence must be parsed enough to identify that these words serve to change the truth-value of the proposition constructed from the other words. It is hard to understand what role “No” is playing in “There are no bears on Mars” until one knows what “bear” and “Mars” mean and how to combine them.

In Experiments 1 and 2, like Austin, et al., we test children’s comprehension of “Not” as a logical operator in the context of a hiding game. A ball is hidden in one of two containers: a truck or a bucket. Children are given verbal information about where the ball is hidden, either in an affirmative form – “It’s in the bucket”; “It’s in the truck” – or with a negative – “It’s not in the bucket”; “It’s not in the truck”. In Experiment 3 we test children’s comprehension of the logical meaning of the word “No”. In Experiment 4 we control for a plausible deflationary account: that the increasing comprehension success for “No” and “Not” reflects increasing inhibitory control, necessary for inhibiting a tendency to go to the container that had been named.

3.2. Experiment 1: Comprehending “Not”

3.2.1. Method

3.2.1.1. Participants

The participants were recruited in separate age groups, with 22 20-month-olds (mean age 20.7, range: 19.2-22.0, 11 boys), 24 24-month-olds (mean 23.8, range: 22.1-25.6, 10 boys) and 26 27-month-olds (mean 27.0, range: 26.1-28.0, 11 boys). All participants were monolingual English speaking children, receiving less than ten hours per week of input in any other language. Thirty-six participants were recruited at the Boston Children’s Museum and participated during their visit. The remaining participants were recruited by phone and email from the greater Boston area and were tested at the Laboratory for Developmental Studies at Harvard University.

Participants from the Boston Children’s Museum were given a sticker for participating.

Participants tested at the Laboratory for Developmental Studies were given a small gift, and their parents were compensated \$5.00 for travel expenses. An additional 19 children (seven 20-month-olds; three 24-month-olds; nine 27-month-olds) were excluded from analysis: nine (three 20-month-olds; no 24-month-olds; six 27-month-olds) for searching in the wrong location on any of the practice trials (see Sec. 3.2.2. below for discussion of this criterion), six (two 20-month-olds; three 24-month-olds; and one 27-month-old) for failure to complete the study, and four (two 20-month-olds; no 24-month-olds; two 27-month-olds) for searching in the same container on every trial (i.e. side bias).

3.2.1.2. Materials

The stimuli consisted of a small 2.5’ tall table, one bright green opaque bucket, one red-and-yellow toy dump truck, a small yellow ball, and a wide black cardboard occluder. Both of the containers, the bucket and the truck, were lined with felt to ensure that the ball would not

make a sound when placed inside. These two containers were chosen because “Bucket” and “Truck” are the earliest known nouns that could readily name containers for the ball (Dale & Fenson, 1996). A ‘jingle box’ was used as a reward activity for successfully finding the ball – the child inserted a ball into an opening on one end, which would roll down a xylophone hidden inside the box, making a sound, and come out the other end.

We also collected productive vocabulary data using the Macarthur Bates CDI short form (Fenson, Pethick, Renda, Cox, Dale, & Reznick, 2000). We modified this form to include the additional words, “Not”, “Bucket”, and “Truck”. Parents were instructed to indicate the words that their children produced, and completed this form before the start of the experiment.

3.2.1.3 Procedure

We created four lists, which counterbalanced the following variables across participants: the container used for the first practice trial (bucket or truck); which side of the table the bucket and the truck were on; and whether the first correct response would be the bucket or the truck. Within participants, the order of experimental trials was always A-B-B-A-A-B-B-A where A and B indicate where the ball was hidden (either the bucket or the truck). The first block of A-B-B-A trials had all negative trials and the second was all affirmative. Since we were interested in the earliest age of success on negative trials and reasoned that doing affirmative trials first may cause children to perseverate onto negative trials, hurting their performance on the latter, we did not counterbalance block order in this experiment (see Experiment 2).

The experimenter stood behind the table and the child sat in their parent’s lap on the floor, 5 feet in front of the table. The table was tall enough that the child could not see into the bucket or truck when they were placed on the table, though they could reach up and touch either container. The experimenter introduced the game to the child as follows: “We’re going to play a

fun hiding game. I'm going to hide this ball, and you're going to find it! When you find the ball, you get to put it in the jingle box!" The experimenter then demonstrated how the jingle box worked. Then the experimenter hid the ball on three practice trials and eight experimental trials. On every trial, after giving the child the clue, the experimenter invited the child to search by saying, "Can you find it? Can you find the ball?" Parents were instructed to encourage the child to find the ball if the child was shy, but without naming or pointing to either container. They were also instructed to release the child to search only once the experimenter asked the child to find the ball.

3.2.1.3.1. Practice Trials

Three practice trials were used to introduce the game. The first practice trial was meant to familiarize children with the occlusion of the containers. In this trial, the experimenter placed only one container on the table: either the bucket or the truck, counterbalanced across subjects. The experimenter then placed the occluder in front of the container and hid the ball. Finally, the experimenter removed the occluder and gave an affirmative clue (e.g., "It's in the truck"). The second and third practice trials were meant to get the child used to making a choice between the two containers, given both an affirmative and a negative clue. In both the second and third practice trials, the experimenter placed both the bucket and truck on the table, in the same positions that would later be used for the test trials. For these two trials, the experimenter hid the ball without using the occluder, so that the child could see where the ball was hidden. In the second practice trial, the experimenter always hid the ball in the other container than the first practice trial and gave an affirmative clue (e.g., "It's in the bucket"). In the third practice trial, the experimenter always hid the ball in the same container as the second practice trial, but gave a negative clue (e.g., "It's not in the truck"). On each practice trial, if a child answered incorrectly

or refused to search, the same trial would be repeated and children would be given a second try. If children failed the second try, the trial would not be repeated again.

3.2.1.3.2. Test Trials

Eight test trials followed the practice trials. The bucket and the truck stayed on the same sides of the table as in the second and third practice trials. On each test trial, the experimenter placed the occluder at the front of the table so that the bucket and truck were no longer visible to the child. The experimenter held up the ball with both hands and said, “Ready? Watch where it’s going!” The experimenter then brought the ball straight down behind the occluder. While looking down at the center of the table, she separated her hands behind the occluder and gently placed the ball in either the bucket or truck. The experimenter touched the bottoms of both containers when placing the ball, to equate any noise resulting from the placement of the ball between the two containers.

For the first four test trials (‘negative trials’), the experimenter told the child where the ball was not, by saying, “It’s not in the [bucket/truck]”. For the last four test trials (‘affirmative trials’), the experimenter told the child the location of the ball by saying, “It’s in the [bucket/truck]”. After each clue, the experimenter invited the child to find the ball by saying, “Can you find it? Can you find the ball?” The experimenter gave affirmative and negative clues with the same friendly prosody, always looking directly at the child, and gave no additional clues through body language.

The child’s choice was taken to be the first object they touched or pointed to. If the child did not touch either the bucket or truck (for example, if they stood in front of a container but never touched it or pointed to it), the trial was excluded. If the child correctly found the ball, he/she got to place it in the jingle box before the next trial. If the child chose the wrong

container, they were shown that the container was empty and told, “It’s not in there”. They were then shown that the ball was in the other container and told, “Look! There it is!” They were not given the ball or allowed to put it in the jinglebox, but were instructed to go back to their parent’s lap and try again. This was meant to encourage the child to try and get it right on the first try.

3.2.2. Results

Children’s choice of the bucket or the truck was recorded during the study by the experimenter. Individual trials were excluded for several reasons. Across all age groups, 17 trials were excluded because the child did not make a clear response (e.g. failed to approach the table, stood in front of the bucket or the truck but did not touch them or indicate a preference, etc.) One additional trial was excluded because the parent gave away the answer before the child searched. For all nonparametric analyses comparing age groups below, children’s performance on affirmative and negative trials was transformed into percentages by dividing the number of trials on which they chose the correct container by the total number of trials of each type (affirmative and negative), after these exclusions. For the logistic regression models, excluded trials were simply treated as missing data points.¹⁵

Before turning to the analysis of the data, a word about the exclusion criteria used in this experiment. In principle, we want to include only those children who understood the hiding game and were trying to find the ball. To avoid the problem of filtering on the same dependent variable as the one being analyzed (see Vul & Kanwisher, 2010), we chose not to exclude any children on the basis of their performance on the eight test trials for any reason other than a 100% side bias. Instead, we took children’s performance on the three practice trials as indicative of their

¹⁵ Unlike Austin, et al. (2014), we do not apply a Bonferroni correction to our results. Each test we perform is theoretical motivated, and given that we expect performance to improve with age, none of the tests are truly independent of each other. Therefore, it is not clear that the probability of a false positive across multiple tests is multiplicative and that a correction is appropriate.

understanding of the game. If children understand that their job is to find the ball, they should at least search correctly when they see exactly where the ball has been placed, as they do in all three practice trials.

Nevertheless, there is no single correct exclusion criterion to apply to the practice trials. Children got two tries on practice trials. Many children were too shy to search on the first try of one or more practice trials. Some children searched in the wrong container on their first try but then searched correctly on the second try – were these children coming to an understanding of the task, or simply trying all the containers? What if they got the next practice trial right from the first try, or, conversely if they got it wrong? It is easy to over-interpret any pattern of choices on three practice trials. Since any exclusion criterion involves some degree of arbitrariness, we analyzed the data using three different, plausible criteria.

We report one consistent exclusion criterion for all experiments in the main analysis below. This criterion excludes any child who searched in the wrong container on any of the practice trials, even on their first try, but does not exclude children who were too shy to search and therefore required multiple tries to make a choice. In this experiment, we analyzed all of the results with two additional exclusion criteria: excluding children who on the first try of more than one practice trial either searched incorrectly or were too shy, and not excluding anyone based on practice trial performance at all. Both of these criteria excluded fewer children. However, we found that while the coefficients in regression models and the values of test statistics changed slightly depending on the choice of the exclusion criterion, no test of significance moved across the threshold of $p=0.05$ in either direction. However, in the other experiments reported below, the choice of exclusion criterion does have some impact on the results involving block order (see Sec 3.2.4. and 4.2.1.4.).

3.2.2.1. Performance by age group

Figure 3.2 shows children's performance in Experiment 1 by age group. To identify the age at which children successfully begin to comprehend the truth-functional meaning of the word "Not", we compared performance within each age group to chance. Children's choices in response to affirmative and negative clues were analyzed separately, using a Mann-Whitney test relative to a chance baseline of 50% correct responding.¹⁶ On negative trials, 20-month-olds chose the correct container significantly below chance ($W=24$, $p=0.02$), 24-month-olds performed at chance levels ($W=114$, $p=0.21$), and 27-month-olds were significantly above chance ($W=106.5$, $p=0.005$). On affirmative trials, 20-month-olds were marginally better than chance ($W=100.5$, $p=0.09$), while 24-month-olds ($W=170$, $p=0.0001$) and 27-month-olds ($W=184$, $p=0.0002$) were highly significantly better (see Fig. 3.2). Coding 20-month-olds' choices to look at performance relative to the mentioned container rather than the correct choice, we find no difference between children's choices on affirmative and negative trials ($W=251$, $p=0.83$). Thus, 20-month-olds chose the bucket equally often whether they heard, "It's not in the bucket" or, "It's in the bucket".

3.2.2.2. Improvement with age

On both types of trials, children's performance improved with age. A Kruskal-Wallis test revealed significant differences in the rates of correct responding between the three age groups on affirmative ($\chi^2=12.11$, $p=0.002$) and negative ($\chi^2=14.29$, $p=0.0008$) trials. According to a Mann-Whitney test, comparing age groups on both types of responses, the youngest age group of

¹⁶ Not counting those children for whom one or more trials were excluded, children could have only gotten 0, 1, 2, 3 or 4 trials correct, in each trial type. This does not allow for a normal distribution of responses, particularly as older children were nearer to ceiling and younger children nearer to floor. In contrast to Austin, et al. (2014), we therefore report only more conservative nonparametric tests for overall group comparisons, and logistic regressions with a binary outcome on each trial (1 for successful search and 0 for failure) for looking at continuous predictors (see below).

20-month-olds had significantly lower rates of correct responding on both affirmative ($W=155.5$, $p=0.01$) and negative ($W=142$, $p=0.006$) trials relative to the middle age group of 24-month-olds. However, the 24-month-olds did not significantly differ from the 27-month-olds on either affirmative ($W=204$, $p=0.34$) or negative ($W=172.5$, $p=0.1$) trials.

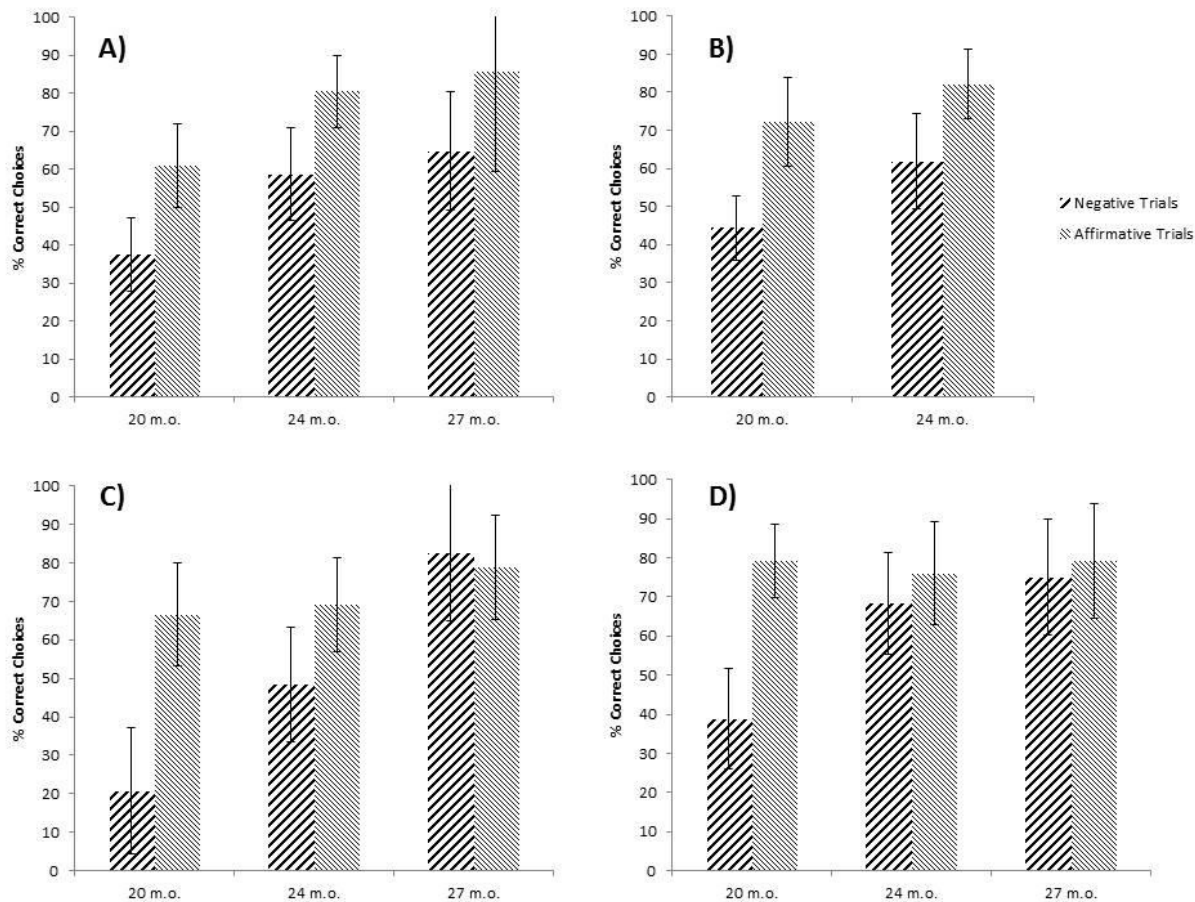


Fig. 3.2: Children's performance on affirmative and negative trials, broken down by age group, experiment, and block order. Error bars indicate 95% confidence intervals. A) Experiment 1, comprehension of logical "Not" with negative trials first; B) Experiment 2, comprehension of logical "Not" with affirmative trials first; C) Experiment 3, comprehension of logical "No" with negative trials first; D) Experiment 3, comprehension of logical "No" with affirmative trials first

We also look at children's improving performance with age, treating age as a continuous variable. Taking each participant's exact age (converting days to fractions of 30-day months), we fit a series of logistic mixed-effect models with grand-mean centered Age as a fixed effect and a random effect of subject with random intercept. Although we are interested in responses to

negative and affirmative trials separately, fitting a model that includes Trial Type as a fixed effect and including data from both types of trials allows us to fully account for between-subject variance by providing more data for the random effect of subject than would fitting a separate models for each Trial Type. To look at performance on negative and affirmative trials separately, we run the model twice, dummy coding either the affirmative or the negative Trial Type as the reference level.

For this and all subsequent experiments, we first fit a model with a three-way interaction between the variables of Trial Type (Affirmative or Negative), Trial Order (1-8), and the continuous variable of Age. The dependent variable is subjects' binary score on every trial (1 or 0). We included the effect of Trial Order because we were interested in the potential presence of practice effects over the course of the experiment, as children became familiar with the task. There was no significant effect of Trial Order and no significant interactions with Trial Order, on performance in either affirmative or negative trials in any of the experiments. This led us to remove this variable from subsequent models, and we do not discuss it further here.

Subsequent models containing only Trial Type and Age showed a significant simple effect of Age on affirmative trials ($\beta=1.13$, 95% CI [1.02, 1.25], $z=2.36$, $p=0.02$)¹⁷ and a highly significant simple effect of age ($\beta=1.22$, 95% CI [1.11, 1.34], $z=4.12$, $p<0.0001$) on negative trials, indicating children's improvement with age.

3.2.2.3. Comprehension performance relative to production of the word "Not"

¹⁷ All β coefficients reported are odds ratios of effect size (that is, exponentiated log-odds coefficients taken from the output of the mixed effects models). We also report 95% confidence intervals for these estimates of effect size.

Using the same logistic mixed-effect modeling approach as above, we look at children's performance as a factor of their production of the word "Not" (once again, with subject as random factor with random intercept), as indicated on the CDI. We find that children's production of "Not" significantly predicts their performance on negative trials ($\beta=3.28$, 95% CI [1.9, 5.7], $z=4.29$, $p<0.0001$), but not on affirmative trials ($\beta=1.49$, 95% CI [0.82, 2.7], $z=1.3$, $p=0.2$). However, it is difficult to separate this effect from the effect of total CDI score. When we fit a model that includes both CDI score (range: 0-103) and production of "Not" as predictors of performance on negative trials, CDI score has a significant independent effect ($\beta=1.04$, 95% CI [1.02, 1.05], $z=5.12$, $p<0.0001$), while there is no independent effect of producing "Not". This is not surprising, given that the production of "Not" is a binary predictor, with parents indicating only whether their child produces it, while the continuous nature of total CDI score makes it possible for this variable to predict more powerfully. Given the limited statistical predictive power of a binary variable, it is nevertheless noteworthy that children's production of the word "Not" is so highly predictive of performance on negative, but not affirmative trials.

3.2.3. Discussion

Testing the comprehension of the logical meaning of "Not", we find that 20-month-olds perform significantly worse than chance, 24-month-olds are at chance, and 27-month-olds do reliably better than chance, with a significant improvement across the entire age range. We find that children's increasing comprehension of logical "Not" tracks closely with their production of that word (although we cannot, at present, distinguish the independent contributions of the increasing production of "Not" from increasing vocabulary and linguistic development more broadly).

Although children's performance on affirmative trials improves with age as well, even the youngest 20-month-old group did marginally better than chance on these trials, indicating that they understand the search task and can use the name of the container to guide their search. When 20-month-olds hear the negative, "It's not in the bucket", they are more likely than chance to go to the bucket. In fact, they are as likely to go to the bucket in this case as when they hear the affirmative clue, "It's in the bucket". This is exactly the pattern of results we would expect if children at this age do not know the word "Not" at all – they treat it as if they had not heard it and simply search in the named container.

In Experiment 1 negative trials always preceded affirmative trials. Although we were concerned about children perseverating from affirmative to negative trials, with their performance on the latter suffering, it is also possible that hearing affirmative sentences can facilitate subsequent processing of similar negative sentences. Since the negative sentences are the same as the affirmatives, with a negation inserted, perhaps some of the processing load of constructing a sentence with a negation can be alleviated by repeatedly constructing the affirmative counterpart first. And perhaps a series of similar affirmative assertions lend pragmatic support to a subsequent negative, making it more natural and easier to process. In Experiment 2 we look for the presence of an order effect between affirmative and negative trials by giving children the affirmative block of trials first. In Experiment 2, we also gave children an extra search attempt, up to three, on each practice trial, if they were either too shy to search or searched incorrectly. We reasoned that this may facilitate their understanding of the experimental paradigm and facilitate performance on both affirmative and negative trials. Experiment 2 also provides the opportunity to replicate the findings of Experiment 1, looking at the age when an understanding of verbal negation develops.

3.3. Experiment 2: Manipulating block order

3.3.1. Method

3.3.1.1. Participants

An additional 43 monolingual English-speaking participants were recruited in the greater Boston area. All of them were tested at the Laboratory for Developmental Studies at Harvard University. None participated in any of the other experiments reported here. The participants were 18 20-month-olds (mean age 20.3, 19.1-21.7, 10 boys), 25 24-month-olds (mean 23.6 months, 22.2-25.6, 10 boys). Recruitment and compensation were identical to Experiment 1. 26 additional toddlers were excluded from analysis: 13 (ten 20-month-olds and three 24-month-olds) were excluded for searching incorrectly on any of the three practice trials, six (four 20-month-olds and two 24-month-olds) for failure to complete the study, and seven (four 20-month-olds and three 24-month-olds) for searching in the same container on all trials (i.e. side bias).

3.3.1.2. Materials and Procedure

Experiment 2 was identical to Experiment 1 in all respects, except that in Experiment 2 participants were given the block of affirmative trials before the block of negative trials.

3.3.2. Results

Across both age groups, only one trial was excluded because the child did not make a clear response (i.e. failed to approach the table).

3.3.2.1. Performance by age group

Figure 3.2 shows the results of Experiment 2, broken down by age group. We followed the same analysis strategy as in Experiment 1, looking at performance on affirmative and negative trials separately. As in Experiment 1, we first look at performance split up into two age groups: 20- and 24-month-olds – corresponding to the youngest and middle age groups in

Experiment 1 (see Fig. 3.2). Unlike the below-chance performance of 20-month-olds in Experiment 1 on negative trials, in Experiment 2 this group was not significantly different from chance ($W=7$, $p=0.24$), 24-month-olds, who performed at chance levels in Experiment 1, did significantly better than chance in Experiment 2 ($W=196$, $p=0.02$). On affirmative trials, 20-month-olds were marginally better than chance ($W=96$, $p=0.054$), while 24-month-olds did highly significantly better than chance ($W=228.5$, $p<0.0001$).

Unlike in other experiments, while performance on affirmative trials does not depend on the choice of exclusion criterion, performance on negative trials does. Using a different criterion and excluding children who failed their first try on multiple practice trials, we would exclude six 20-month-olds and no 24-month-olds. In this case, 24-month-olds still perform better than chance ($W=76$, $p=0.03$), but 20-month-olds are now significantly below chance ($W=10$, $p=0.03$). Applying no exclusion criterion at all, the total sample would include 28 20-month-olds and 28 24-month-olds, with 24-month-olds performing marginally above chance ($W=212$, $p=0.07$), while 20-month-olds perform significantly worse than chance ($W=10$, $p=0.03$). We thus conclude that while 24-month-olds perform consistently better than chance in Experiment 2 (unlike Experiment 1), whether 20-month-olds are significantly below chance (as they were in Experiment 1) depends the exclusion criterion, and likely on sample size – the fewer children excluded, the more robust the significant performance below chance on negative trials in this group.

3.3.2.2. Improvement with age

Mann-Whitney tests comparing the two age groups on both types of responses revealed a significant difference between the 20-month-olds and the 24-month-olds on both affirmative ($W=151$, $p=0.047$) and negative ($W=123.5$, $p=0.01$) trials. As in Experiment 1, taking each

participant's exact age (converting days to fractions of 30-day months), we fit a series of logistic mixed-effect models with grand-mean centered Age as a fixed effect and a random effect of subject with a random intercept. Models containing only Trial Type and Age as fixed effects showed a significant simple effect of Age on affirmative trials ($\beta=1.34$, 95% CI [1.07, 1.71], $z=2.47$, $p=0.01$) and on negative trials ($\beta=1.23$, 95% CI [1.02, 1.5], $z=2.1$, $p=0.04$). Note that since Experiment 1 had an additional group of older children included in the analyses, the full age ranges from the two experiments are not directly comparable.

3.3.2.3. Comprehension performance relative to production of the word “Not”

As in Experiment 1, we use the same modeling strategy to look at whether production of the word “Not” is a significant predictor of performance. We do not find a significant effect of producing “Not”, as indicated on the CDI, on performance on either affirmative or negative trials. We attribute this to a lack of statistical power, due to the smaller sample and more limited range of age (and performance) in Experiment 2, relative to Experiment 1.

3.3.2.4. Block order effects: Comparison of Experiments 1 and 2

Given that 24-month-olds in Experiment 2 perform better than chance on negative trials, while the same age group in Experiment 1 did not, we compare similar age groups from the two experiments directly in order to look for an effect of block order. Mann-Whitney tests comparing the youngest 20-month-old groups in Experiments 1 and 2 find no significant difference between the two on either affirmative ($W=150$, $p=0.18$) or negative ($W=150$, $p=0.17$) trials. Comparing the 24-month-old groups from the two experiments also shows no significant differences on either type of trial (affirmative: $W=295.5$, $p=0.36$; negative: $W=246$, $p=0.27$). Taking both age groups and building a logistic mixed-effects model with a random intercept for participants and Experiment (1 or 2) as a fixed effect, we find no significant effect of this variable on

performance in either the affirmative or negative trials. While the performance of 24-month-olds relative to chance depends on the exclusion criterion, the lack of significant difference between performance on Experiments 1 and 2 does not.

3.3.3. Discussion

In Experiment 1, when doing negative trials before affirmative, 20-month-olds performed below chance and 24-month-olds were at chance on negative trials. In Experiment 2, 20-month-olds stay below chance (whether significantly so depends on the choice of exclusion criterion), and 24-month-olds perform significantly better than chance (regardless of exclusion criterion). Although there was no difference between the performance of 24-month-olds in Experiments 1 and 2, the finding that they are at chance on the former and above-chance on the latter suggests that receiving affirmative trials before negative trials may help 24-month-olds succeed in understanding the word “Not”. This pattern suggests that 20-month-olds do not, on average, comprehend the logical meaning of “Not”, whether they get practice and pragmatic support by first hearing highly similar affirmative clues. On the other hand, 24-month-olds may be at a transitional age at which building representations of affirmative propositions facilitates later processing of those propositions combined with a logical negation operator.

Taken together, Experiments 1 and 2 show that the ability to comprehend the word “Not” as logical negation emerges early in the third year of life. This tracks quite closely with children’s emerging production of the word “Not” (Dale & Fenson, 1996; see Fig. 3.1 for population norms and Fig. 3.3 for production data from the present sample). At 27 months – the average age of the group of children who succeed in Experiment 1 – just over 50% of children are reported to produce this word in normed CDI data. Our findings replicate the age of comprehension success of logical “Not” that had been identified by Austin, et al. (2014). In

contrast to “Not”, “No” is one of the earliest words children produce, with 55% of 15-month-olds producing “No”, at an age when there are almost no reports of any child producing “Not”. Austin, et al. find that children comprehend “No” as a logical operator at the same age as “Not”. In Experiment 3 we seek to replicate this pattern, while again relating each child’s comprehension success to their reported production of both “No” and “Not”.

We also look at children’s performance across all three experiments on affirmative relative to negative trials. Affirmative trials might be easier for one or both of the following reasons: negative statements might be more difficult to comprehend than affirmatives (if comprehending sentences containing logical “No” and “Not” is pragmatically or semantically difficult); and, even once a negative statement has been well-understood, the decision-making process involved in searching for the ball may be easier given affirmative than negative clues. Once a child has comprehended an affirmative clue, the decision of where to search follows. To find the ball from a negative clue, the child must not only comprehend where the ball is not, but make an additional inference to where it might be. Whether one or both factors play a role, we expect children to perform better on affirmative than negative trials, particularly at younger ages, when their performance on negative trials is at or below chance. Yet Austin, et al. (2014) find chance performance on affirmative trials even at older ages, when success on negative trials is reliable. Though they do not compare affirmative to negative trials directly, the finding that children are at least not clearly better on affirmative trials is surprising.

3.4. Experiment 3: Comprehending “No”

To adapt the paradigm of Experiments 1 and 2 to include No, we introduced a question-answer interaction between the parent and the experimenter, more similar to Austin, et al. (2014). The parent, sitting with the child and unable to see where the ball is hidden, asks the

experimenter, Is it in the bucket/truck? The experimenter would answer, “Yes, it is”. Or “No, it’s not”. As before, the child would be invited to search for the ball. Although the negative clue in this construction includes the word “Not”, we have already determined in Experiments 1 and 2 that children below either 24 or 27 months (depending on the block order) do not comprehend “Not”. In this experiment, we are interested primarily in the success of younger children, so that the inclusion of “Not” in the negative clue is unlikely to help them, except to make the utterance sound more natural and less abrupt.

3.4.1. Methods

3.4.1.1. Participants

Seventy monolingual English-speaking participants were recruited in the greater Boston area, across the same age groups as in Experiment 1. All of them were tested at the Laboratory for Developmental Studies at Harvard University and none participated in Experiments 1 or 2 (one child had previously participated in Experiment 4). The participants were 20 20-month-olds (mean age: 20.6, 19.2-22.0, 11 boys), 25 24-month-olds (mean age: 23.4, 22.2-, 11 boys) and 25 27-month-olds (mean age: 27.3, 26.2-29.0, 13 boys). Recruitment and compensation were identical to Experiment 1. An additional 34 children (17 20-month-olds; 12 24-month-olds; four 27-month-olds) were excluded from analysis: 11 (four 20-month-olds; seven 24-month-olds; no 27-month-olds) for searching in the wrong location on any of the practice trials, 14 (seven 20-month-olds; three 24-month-olds; and four 27-month-olds) for failure to complete the study, and six (four 20-month-olds; two 24-month-olds; no 27-month-olds) for searching in the same container on every trial (i.e. side bias). Two additional 20-month-olds were excluded because of parent error (parents reading the wrong line in their script and asking the wrong question on one of the trials).

3.4.1.2. Materials and Procedure

The procedure differed from Experiments 1 and 2 only in the clues given and how these involved interacting with the parent. Parents were given a script to follow during the experiment. On every trial, parents waited for the experimenter to finish hiding the ball and remove the occluder. The experimenter then made eye contact with the parent, which served as a cue for the parent to turn to the child and say a phrase to engage them: Hm, where did it go? or, I wonder where it is! Then the parent would look at the experimenter and ask, Is it in the bucket? or, Is it in the truck? depending on the trial. The script instructed the parent which container to ask about, and the order of mention of the containers was counterbalanced in an ABBA pattern, the same as in Experiments 1 and 2. The parent was not aware of the actual location of the ball until the experimenter answered their question.

Just as in Experiments 1 and 2, affirmative and negative trials were blocked, so that all four trials of the same type occurred in a row. Experiment 3 counterbalanced block order across subjects, so that half of the children received the affirmative trials first, and half received the negative trials first. In the affirmative trials, the experimenter answered, Yes, it is. In the negative trials, the experimenter answered, “No, it’s not”. Parents’ scripts always corresponded to the experimenters’ answers so that the answers were always correct relative to the ball’s actual location. The experimenter gave affirmative and negative clues with identical prosody, making eye contact directly with the parent, and gave no additional clues through body language. After giving the clue, the experimenter looked at the child and invited them to find the ball. Parents were instructed not to give any additional clues. We excluded from analysis trials in which there was parental interference. For all participants, if the child correctly found the ball, he/she got to

place it in the jingle box before the next trial. If they did not, they returned to the parent to try again.

3.4.2. Results

3.4.2.1. Comprehension of “No”

Across all age groups, seven trials were excluded because the child did not make a clear response (e.g. failed to approach the table, stood in front of the bucket or the truck but did not touch them or indicate a preference, etc.) One additional trial was excluded due to experimenter error and one for parental interference.

3.4.2.1.1. Performance by age group

Figure 3.2 shows the results of Experiment 3, broken down by age group and block order. As in Experiments 1 and 2, we first compare the performance of each age group to chance to look for an age of success. On negative trials, 20-month-olds chose the correct container significantly below chance ($W=16.5$, $p=0.04$), 24-month-olds performed significantly better than chance ($W=147.5$, $p=0.049$), as did 27-month-olds ($W=230$, $p=0.0005$). On affirmative trials, all three age groups performed better than chance (20-month-olds: $W=105$, $p=0.0008$); 24-month-olds: $W=147.5$, $p=0.0006$; 27-month-olds: $W=221$, $p=0.0002$). Coding 20-month-olds' choices by performance relative to the mentioned container rather than the correct container, just as in Experiment 1, we find no difference between children's choices on affirmative and negative trials ($W=144$, $p=0.11$). Thus, when hearing the question, Is it in the bucket? 20-month-olds chose the bucket as often when the answer was “Yes, it is” as “No, it's not”.

The above-chance performance of 24-month-olds depended on the exclusion criterion. When no children were excluded based on the practice trials, and additional three children were included in this age group. Their inclusion made this groups performance no longer different

from chance ($W=202$, $p=0.12$). The fragility of this group's performance relative to chance depends on order effects (See Sec. 3.4.2.1.4. below).

3.4.2.1.2. Improvement with age

Unlike Experiments 1 and 2, children's performance on affirmative trials was not significantly different between the three age groups, according to a Kurskal-Wallis test ($\chi^2=0.87$, $p=0.65$). Children's performance on negative trials, on the other hand, did differ between age groups ($\chi^2=17.75$, $p=0.0001$). Comparing performance on negative trials between pairs of age groups using a Mann-Whitney test, the youngest age group of 20-month-olds had significantly lower rates of correct responding ($W=241$, $p=0.02$) relative to the middle age group of 24-month-olds. The 24-month-olds were significantly worse than the 27-month-olds ($W=220$, $p=0.01$). There were no significant differences between pairs of groups on affirmative trials.

Next, we again fit a logistic mixed-effects model with score on each trial as the binary dependent variable, grand-mean centered Age as a predictor and a random effect of subject with random intercept. These models showed no significant effect of age on affirmative trials ($\beta=1.08$, 95% CI [0.98, 1.2], $z=1.56$, $p=0.12$), but a highly significant effect of age on negative trials ($\beta=1.35$, 95% CI [1.22, 1.5], $z=5.66$, $p<0.0001$).

3.4.2.1.3. Comprehension performance relative to production of “No” and “Not”

Figure 3.3 shows the trajectory of children's comprehension performance with age on both “No” and “Not”, alongside their rate of production of both words. As in Experiments 1 and 2, we use the same modeling approach to look at whether children's performance in these comprehension tasks tracks with their production of the words “No” and “Not”. We find that production of “No” does not significantly predict performance on either affirmative or negative trials. However, production of “Not” significantly predicts performance on negative trials

($\beta=2.17$, 95% CI [1.26, 3.85], $z=2.73$, $p=0.006$), but not on affirmative trials. As in Experiments 1 and 2, however, there is no independent contribution of producing “Not” in a model that also includes total CDI score as a predictor. We conclude that production of “Not”, but not of “No”, tracks with comprehension success of the truth-functional meaning of the word “No”, just as it did for comprehending that meaning of “Not”. This also suggests that the methodological differences between the single-experimenter design of Experiments 1-2 on the one hand and the design of both Experiment 3 and Austin, et al. (2014) on the other, where children overheard the clues as a dialog between two adults, did not impact children’s comprehension performance.

3.4.2.1.4. Block order effects

Figure 3.2 shows the results of Experiment 3 broken down by block order. Since we manipulated block order within Experiment 3, we can directly compare the performance of children who received the affirmative block first to those who received it second. We find that getting affirmative trials first increases success with negative trials only at the transitional age group of 24-month-olds. Although there are half as many children in this comparison as in the equivalent comparison between Experiments 1 and 2, a Mann-Whitney test finds a marginally significant difference at the transitional age for successful performance on negative trials – between the 24-month-old children who got affirmative clues first and those who got negative clues first ($W=42.5$, $p=0.07$). We find no such differences for other age groups. Those 24-month-olds who received the affirmative block first performed significantly better than chance on a Mann-Whitney test ($W=58$, $p=0.02$), while those who received the negative block first performed at chance levels ($W=9.5$, $p=0.91$). This indicates that the overall above-chance performance of 24-month-olds in Experiment 3 is driven by those children who received the

affirmative block of trials first. In contrast, children's performance on affirmative trials did not differ by block order in any age group.

However, as in Experiment 2, the statistical significance of order effects depends on the exclusion criterion. The results reported above hold when applying the most stringent criterion, excluding those children who fail any of the practice trials on the first try. Using no exclusion criterion (which includes an additional three children) reduces the effect, with the performance of 24-month-olds who received the affirmative block of trials becoming marginally different from chance ($W=78$, $p=0.1$), and no difference approaching significance between the children who got affirmative trials first and those who got negative trials first.

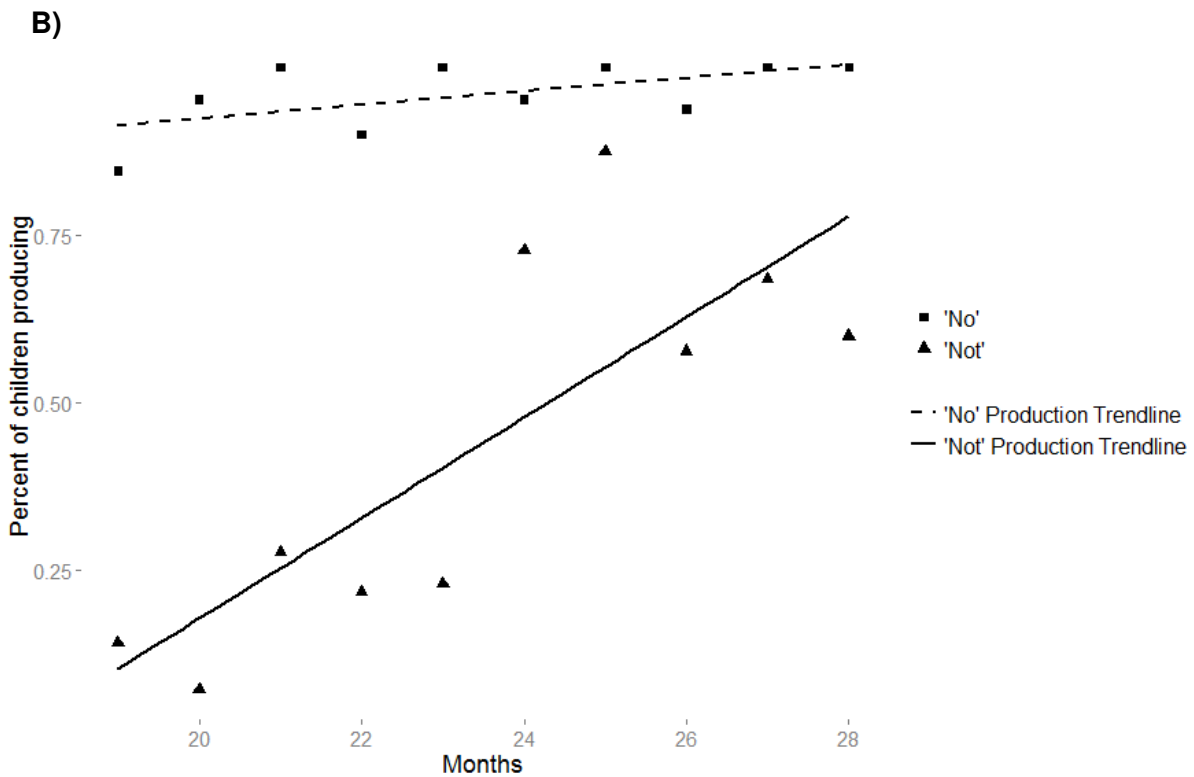
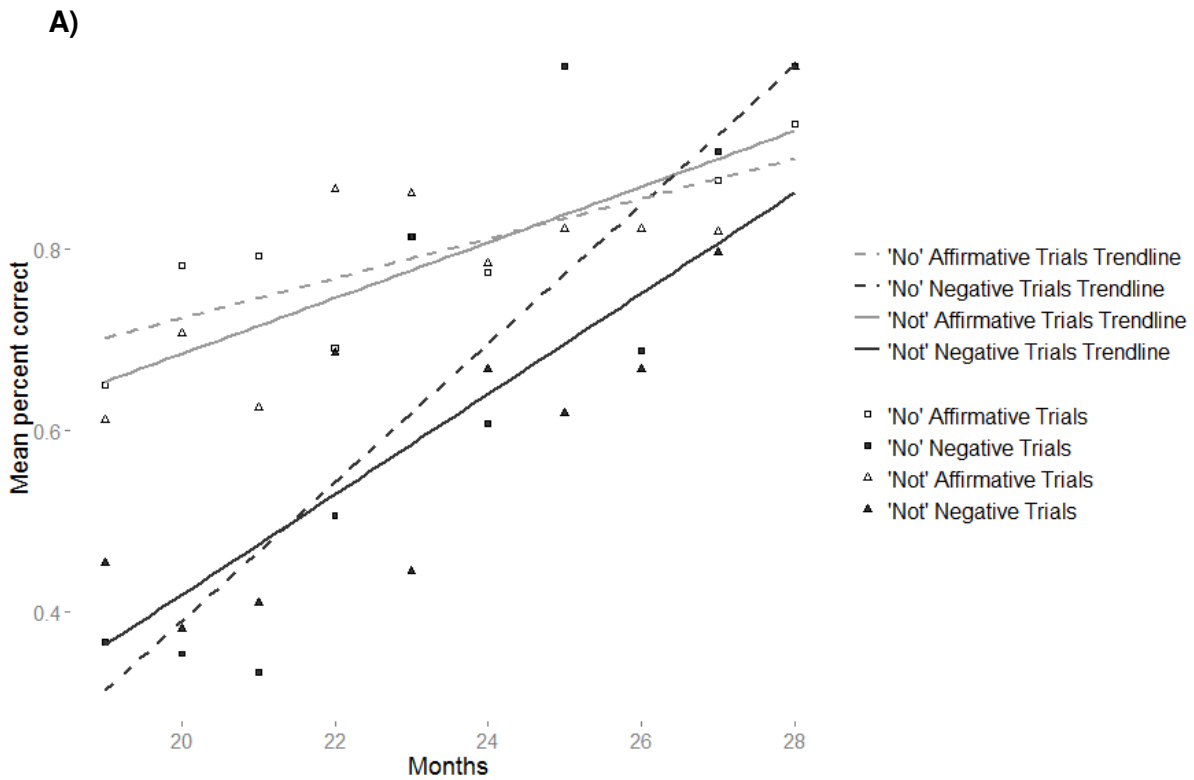


Fig. 3.3: A) Children's performance on affirmative and negative trials in Experiments 1 and 2 (pooled together) and Experiment 3, by month. B) The percent of all children in Experiments 1-3 producing the words "No" and "Not" by month, taken from parents' report on the CDI.

3.4.2.2. Comparing Experiment 3 to 1 and 2: Comprehending logical “No” vs. “Not”

3.4.2.2.1. Comparing ages of success

Experiment 3 confirms the findings of Austin et al., that 2-year-olds begin to comprehend “No” and “Not” at the same time, despite producing “No” at a much younger age. Figure 3.2 shows the comparison of children’s performance in Experiment 3 to Experiments 1 and 2, broken down by age group. Since block order is manipulated between Experiments 1 and 2 and within Experiment 3, for the purposes of comparing Experiments 1 and 2 to Experiment 3, we pool participants from Experiments 1 and 2 together for the 20- and 24-month-old groups. Since we did not test 26-month-olds in Experiment 2, such that all 26-month-olds tested on comprehension of logical “Not” received only the negative trials first, we compare these children to just those 26-month-olds in Experiment 3 who also got the negative block of trials first. There were no significant differences between the results of Experiments 1 and 2 and Experiment 3 in any age group according to a Mann-Whitney test (20-month-olds: $W=251$, $p=0.69$; 24-month-olds: $W=661$, $p=0.57$; 27-month-olds: $W=112$; $p=0.34$).

Looking at age as a continuous variable, we fit logistic regression models with fixed effects for Experiment (1 or 2, vs. 3), Block Order (Affirmative-First vs. Negative-First) and Age, centered by the mean of participants’ ages from all three experiments, as well as the full set of interaction terms between these variables. As in previous models, there was a random effect of participant, fit with a random intercept. The only significant predictor of performance on either affirmative or negative trials was Age (Affirmative: $\beta=1.16$, 95% CI [1.04, 1.31], $z=2.56$, $p=0.01$; Negative: $\beta=1.27$, 95% CI [1.15, 1.41], $z=4.60$, $p<0.0001$). There were no significant interactions, and no other significant fixed effects.

3.4.2.2.2. Comparing affirmative and negative trials

We also compare performance on affirmative and negative trials across all three experiments. As Figure 3.2 shows, performance on affirmative trials is consistently better than performance on negative trials across all three experiments, independently of age ($W=22828.5$, $p<0.0001$). Broken down by age group, this difference holds for 20-month-olds ($W=2802.5$, $p<0.0001$) and 24-month-olds ($W=3811.5$, $p<0.0001$), but not at the oldest group of 27-month-olds ($W=1382$, $p=0.55$), at which point performance on negative trials has caught up (see Fig. 3.2 and 3.3). In a logistic regression looking at the performance of children at all ages across all four experiments and including the variables of Trial Type (Affirmative or Negative), Experiment (1, 2 and 3) and Age, along with all interactions, there were no significant interactions with the variable of Trial Type.

3.4.3. Discussion

Confirming the results of Austin, et al. (2014) on negative trials, Experiment 3 shows that children begin to understand the logical meaning of the word “No” at the same time as they understand the same meaning of “Not”. They are below chance at 20 months, above chance (when getting the affirmative block of trials first) at 24 months and above chance regardless of block order at 27 months. While in Experiments 1 and 2 we found that the developmental trajectories of truth-functional comprehension and production of “Not” track closely together, comprehending truth-functional “No” does not track with production of the word “No” at all. Rather, as Figure 3.3 shows, it tracks with producing “Not”. The comprehension of “No” as a logical operator lags production of the word by about a year. This suggests that mapping the concept to the word poses a significant challenge – either because the concept is not available at 15 months when the word is first learned, or because the mapping problem is a particularly difficult one. The gap also suggests that when younger toddlers produce “No”, they use it with

some meaning other than the logical one. This, in turn, suggests that young children's main early use of "No" – to signify 'rejection' – is indeed likely to rely on a semantically different concept rather than being a different usage of the logical concept, just as Bloom (1970) hypothesized.

3.4.3.1. Converging results

Although Austin, et al. (2014) do not compare comprehension to production in their data, they do provide converging evidence for similar ages of comprehension success. They find success at 24 months, but not earlier, when children are given the single-word clue "No" as well as the word "Not" contained in a sentence like, "It's not in the bucket". Although the clue used in our Experiment 3 is "No, it's not" rather than a single-word "No", that we find success at a similar age range to Austin, et al. suggests that the word "Not" is playing no role in children's responses on negative trials in Experiment 3. However, we find robust success at 27 months in comprehension of logical "No" and "Not" both, with success at 24 months depending on block order, while Austin et al. report success at 24 months with alternating affirmative and negative trials. What accounts for these differences?

First, Austin, et al. exclude 30% of children (53/179) across all ages for failing to complete at least three quarters of the trials, and 29% of children (17/59) at the critical group of 24-month-olds. Using the same criterion, we exclude 12% (26/211) across all ages and 9% (7/74) of the 24-month-olds. While it is unclear what difference between either our methods or samples caused this large difference in exclusion rates, it is reasonable to assume that the same fussy children are more likely to both choose incorrectly and quit the study early, so that excluding more children overall would likely lead to a slightly earlier age of success in Austin, et al.'s experiment. In addition, while Austin, et al. find success on negative trials at 24 months, they do not find success on affirmative trials, using the construction *It's in the bucket/house* either at 24

months or even the oldest group in their experiment, 29 month-olds. Though we do not apply a Bonferroni correction, we find a consistent pattern of success on affirmative trials that would survive this correction for the 24- and 27-month-old age groups, as well as reliably greater success on affirmative relative to negative trials across the three experiments. This pattern of results gives us confidence that we have found a reliable age of success on negative trials within the same paradigm.

Austin, et al. also use a paradigm in which two experimenters are interacting and the child is observing and overhearing their conversation. One experimenter asks whether the ball is in the bucket or the house, and the other responds with, “It’s in the bucket/house” or “It’s not in the bucket/house”. While we use a similar question-answer design in Experiment 3, Experiment 1 and 2 have just one experimenter hiding the ball and then simply telling the child either where it is or is not. Despite all of these differences, both Austin, et al. and the present report converge on the range around 23-27 months as the age at which children begin to comprehend logical negation.

3.4.3.2. Inhibitory control: the deflationary hypothesis

Taken together, the results of the three preceding experiments suggest that there is some common factor, developing between 19 and 28 months, that leads to understanding the logical meanings of both English negation words, “Not” and “No” at the same time. In addition to the conceptual and linguistic accounts we have mentioned, there is another deflationary explanation. Successfully acting on negative information, as given in a sentence like, “It’s not in the bucket” or the question-answer dialogue, “Is it in the bucket?” “–No, it’s not!” requires forming a representation of the bucket before shifting to the truck. Inhibiting attention (and in a search task, inhibiting motor planning and action) towards the referent of the affirmative predicate in order to

attend to its negated counterpart is a necessary aspect of processing all logical negation, and so is common to the processing both of “No” and “Not”. Indeed, Nordmeyer and Frank (2014) find evidence that inhibitory control is a source of great difficulty for children’s online processing of sentences containing negation. They find that children much older than 28 months fail to demonstrate comprehension of logical negation when presented with referents for both the affirmative predicate and its negative counterpart. They argue the failure is due to children’s attention being strongly drawn to the affirmative referent when they hear its name and then getting stuck there, failing to ever shift over to the negative.

Indeed, in our search paradigm, inhibitory control may pose difficulty for children not only in sentence processing but also after, in deciding where to look for the ball, even if they have successfully processed the negative verbal information. Since success on negative trials requires an additional inferential step – from where the ball is not to where it might be – and since this step requires inhibiting attention to the first location the child has considered, a failure of inhibitory control may also explain why young children perform better on affirmative than negative trials, and improving inhibitory control with age may explain the shrinking of this difference, which we found over Experiments 1-3 across the age groups.

Both children’s better performance on affirmative relative to negative trials at younger, but not older ages, and their improving success on negative trials over the range of ages we tested may thus reveal something about their improving inhibitory control and executive functioning, not an improving understanding of negation. It would account not only for the chance behavior of the middle age group of 24-month-olds, but also for the systematically incorrect choices of the youngest 20-month-old group. These children could be choosing the named container, whether the word is in a negative or an affirmative statement, because they

hear its name and cannot shift their attention away from it. Experiment 4 examines this possibility.

3.5. Experiment 4: Testing the inhibitory control hypothesis

In Experiment 4, we place a similar inhibitory demand on children as in Experiments 1-3, while changing the way they are given negative information. Instead of being told that the ball is not in the bucket or in the truck on negative trials, they are shown that the bucket or the truck is empty. To do this, the experimenter looks at the empty container, lifts it up and tilts it to show its contents to the child, while saying, Look at the [bucket/truck]! The other container is not named or highlighted in any way. We test the youngest age group from Experiments 1-3, reasoning that if 20-month-olds fail to comprehend negation because they cannot inhibit their attention to the named container, they ought to fail in this task as well. If, on the other hand, increasing success with age reflects an increasing facility with the language of logical negation, and if even the youngest children can reason from the visible emptiness of a container to searching in a different location, then the youngest children should succeed.

3.5.1. Methods

3.5.1.1. Participants

Twenty-four monolingual English-speaking participants (mean age: 21.0, 19.2-22.7, 12 boys) were recruited in the greater Boston area. All of them were tested at the Laboratory for Developmental Studies at Harvard University and none had previously participated in any of the other experiments reported here. Recruitment and compensation were identical to Experiment 1. Twelve additional toddlers were excluded from analysis: four for searching incorrectly on any of the practice trials, six for failure to complete the study, and two for only searching in one location (i.e. side bias).

3.5.1.2. Materials

The stimuli were identical to those of Experiments 1-3, except that a slightly larger yellow ball was used for the task. This change was made so that the ball would be more easily visible to children on affirmative trials when the container was tipped towards them, without needing to tilt the container so far that the ball would roll out.

3.5.1.3. Procedure

As in Experiments 1 and 2, there was no dialog between the parent and the experimenter in this experiment. The verbal clues in Experiments 1-3 were replaced by visual clues. To give a clue, the experimenter would look at the container, pick it up and tilt it until it was lying sideways on the table and the child could see inside. While tilting it over, the experimenter said, Look at the [bucket/truck]! On affirmative test trials, the ball was visible within the container, and on negative trials, the container was empty (with the ball hidden in the other container). The experimenter kept the container down for approximately two seconds while the child was looking at it. If the child was not looking when the container was first tilted, the experimenter called the child's attention to the container, and held the container down until the child looked, keeping it down for two seconds after the child's attention was drawn. The experimenter then returned the container to its upright position and invited the child to search. Since the last thing the experimenter had told the child was to look at the container, we reasoned that children might be confused by the phrase, Can you find it, can you find the ball? that was used in the previous experiments, thinking It might refer to the container instead of the ball. So in Experiment 4, we reversed this instruction, asking the children, Can you find the ball, can you find it? All other details of counterbalancing were kept the same as in the previous experiments, and block order was counterbalanced as in Experiment 3.

3.5.2. Results

Four trials were excluded because the child did not make a clear response (e.g. failed to approach the table, stood in front of the bucket or the truck but did not touch them or indicate a preference, etc.)

3.5.2.1. Performance relative to chance

Figure 3.4 shows the results of Experiment 4, alongside results from the 20-month-old age group in Experiments 1 and 2 (averaged together), as well as Experiment 3. We find that, on both affirmative and negative trials, 20-month-olds in Experiment 4 chose the correct container significantly above chance (See Fig. 3.2; Affirmative: $W=349$, $p=0.0001$; Negative: $W=141$, $p=0.01$). There were no significant effects of block order on either affirmative or negative trials.

3.5.2.2. Performance relative to Experiments 1-3

Comparing these participants' performance to the same age groups in Experiments 1, 2 and 3, a Kruskal-Wallis test revealed a significant difference between the Experiments on negative ($\chi^2=15.39$, $p=0.002$) and affirmative ($\chi^2=14.51$, $p=0.002$) trials. A series of pairwise Mann-Whitney tests, comparing these participants' performance to each of the previous experiments, shows significant differences between Experiment 4 and Experiment 1 on Affirmative ($W=417.5$, $p=0.0004$) and Negative trials ($W=415$, $p=0.0007$), between Experiment 4 and Experiment 2 on Affirmative ($W=157$, $p=0.03$) and Negative ($W=157$, $p=0.04$) trials. Between Experiment 4 and Experiment 3, the difference is marginal on Affirmative trials ($W=295.5$, $p=0.07$), but significant on Negative trials ($W=348$, $p=0.003$). There were no significant effects of block order on either Affirmative ($W=66.5$, $p=0.97$) or Negative ($W=53$, $p=0.39$) trials.

3.5.2.3. Comparing affirmative and negative trials

As in Experiments 1-3, the 20-month-olds in Experiment 4 performed better on affirmative than negative trials ($W=433$, $p=0.002$). In a logistic regression looking at the performance of the 20-month-olds across all four experiments and including the variables of Trial Type (Affirmative or Negative), Experiment (1, 2, 3, and 4) and their interactions, there was no significant interaction between Trial Type and any of the experiments.

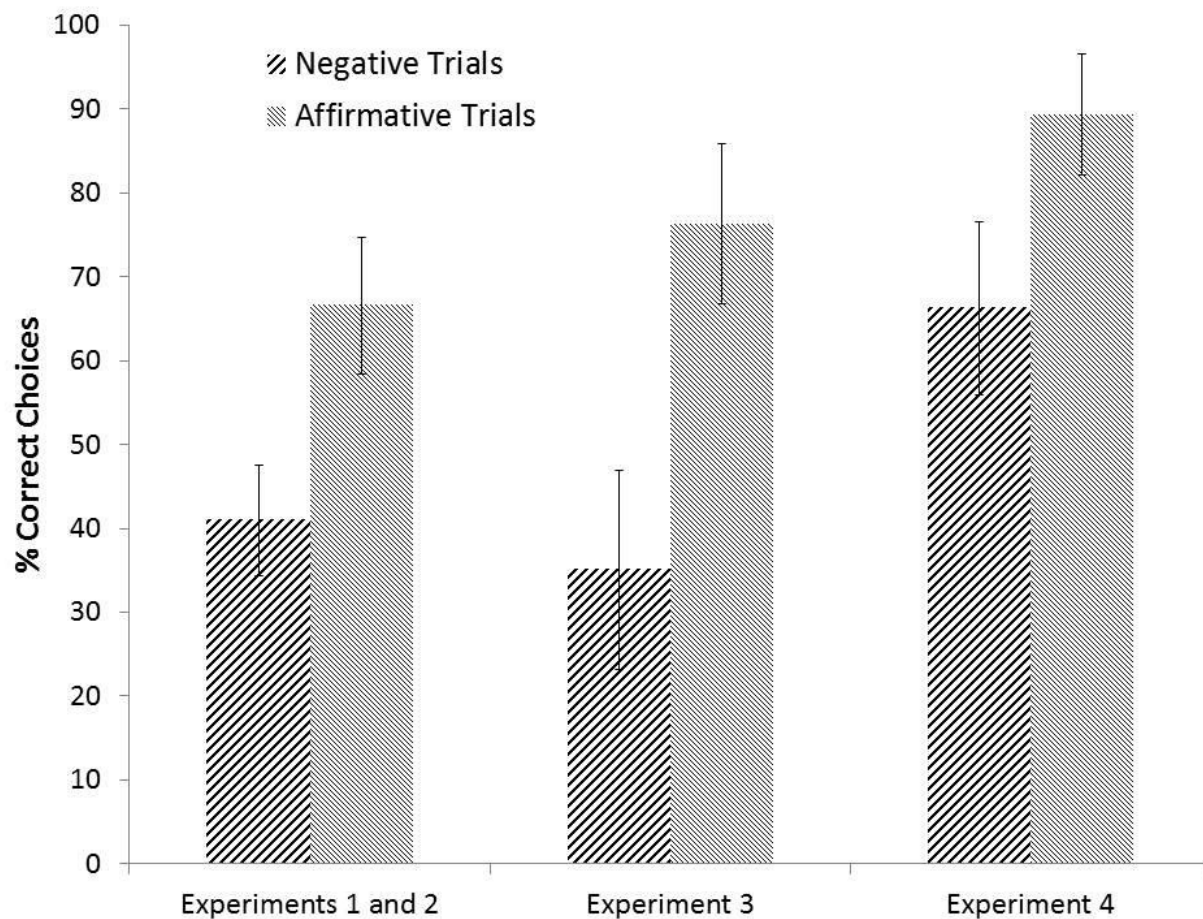


Fig. 3.4: Performance on negative and affirmative trials of the 20-month-old age group in Experiments 1-4. Data from Experiments 1 and 2 is combined, with the mean reflecting average comprehension performance with “Not”, across both block orders. Error bars indicate 95% confidence intervals.

3.5.3. Discussion

In Experiments 1-3 we find that 20-month-old children go to the named container, ignoring both the words “No” and “Not” to search in the wrong location. In Experiment 4 we

find that children at this age can nevertheless successfully avoid a container that they have seen is empty to find the ball at above-chance levels, even when the wrong container is named and children's attention is drawn to it. That they succeed in this case suggests that immature inhibitory control alone cannot be responsible for their failure in Experiments 1-3, nor for the increasing success of children with increasing age. If anything, the inhibitory control needed to succeed in Experiment 4 is greater than in the preceding experiments. While in Experiments 1-3 children hear the name of the wrong search location when given the negative clue, in Experiment 4 they not only hear the wrong container named, but that container is also attended to by the experimenter, picked up, and tilted down so they can see into it.

While the inhibitory demands of Experiment 4 are greater, one might argue that the information provided children is of higher quality. Even if children did understand the meaning of "It's not in the bucket" or "Is it in the bucket?" "–No, it's not", seeing that the bucket is empty for themselves may be better evidence of where the ball is than hearing about it from the experimenter. Indeed, evidence that children do better with visual than verbal information more generally comes from their greater success on affirmative trials in Experiment 4 relative to Experiments 1-3. Although the advantage of visual over verbal information likely does play some role in children's improved performance in the negative trials of Experiment 4, it cannot fully explain their success. If children understood the negative verbal clues given in Experiments 1-3, but either found the information unreliable or did not encode it as well, they would have performed at chance. Instead, 20-month-olds performed significantly below chance when given verbal negative clues, indicating that they did use the information they were given – incorrectly. They reliably used the words Bucket and Truck to guide their search. They just did not use "No" or "Not".

The results of Experiment 4 also bear on interpreting the difference in performance between affirmative and negative trials in all four experiments. Better performance on affirmative than negative trials is over-determined. It could reflect difficulties in processing negation words in particular, or in inhibiting attention to the affirmative referent more generally. Twenty-month-olds in Experiment 4 perform better on affirmative than negative trials by the same margin as in Experiments 1-3, despite no verbal negation being used in Experiment 4. This suggests that the constant size of this performance difference reflects the inhibitory demands – avoiding going to the named container – that are common to all four experiments. Note that this does not mean that the comprehension of negation words poses no separate difficulty – in fact, it must to account for 20-month-olds below-chance performance with both verbal “No” and “Not”. It is just that this difficulty is not additive on top of the significant difficulty already posed by the inhibitory demands of negative trials. The advantage of affirmative over negative trials in Experiment 4 supports the idea that the difference between the two types of trials in Experiments 1-3 is not a good indicator of the difficulty of verbal negation in particular.

Although children’s inhibitory control is no doubt developing between 19 and 28 months, the results of Experiment 4 show that this development cannot completely account for their increasing success in comprehending negation words as logical operators. What representation guides children’s successful search in Experiment 4? One possibility is that children possess a concept of logical negation and use it to represent that the ball is not in the bucket when they see an empty bucket, but have not mapped this concept to either the words “No” or “Not”. The other possibility is that children do not yet have a concept of logical negation, and are relying on a different concept to succeed on the negative trials in Experiment 4, such as a concept of

emptiness in relation to containers, and a heuristic that guides them to avoid searching in places known to be empty.

3.6. General Discussion

In Experiments 1 and 2 we find that understanding the logical force of the word “Not” parallels its emerging production, with 20-month-olds ignoring this word entirely, 24-month-olds successfully using it to guide their search only if they have done a block of affirmative trials first, and 27-month-olds showing robust comprehension of the logical meaning of “Not” regardless of block order. In Experiment 3 we find that although “No” is produced much earlier than “Not”, the comprehension of truth-functional “No” emerges at just the same age as the comprehension of truth-functional “Not” and tracks with production of “Not” rather than “No”. That the comprehension of logical “No” and “Not” track so close to each other (see Fig. 3.3) suggests that a single common factor underlies children’s ability to map both words to the concept of negation. Experiment 4 rules out one deflationary account of that factor, showing that it cannot be increasing inhibitory control relative to the named container (the wrong choice on negative trials), since even the youngest children can make the correct choice when their attention is drawn to a container that is shown to be empty.

3.6.1. Conceptual versus linguistic factors

What is the common factor driving the similar developmental course of comprehending both “No” and “Not” as logical operators? We see two possibilities. One is that the limiting factor is conceptual – children younger than 24 months on average do not have the concept of logical negation. When they acquire it early in the third year of life, they quickly map both of the words “No” and “Not” to it. The other possibility is that the limiting factor is linguistic. On this account children do have the concept of a logical negation operator, but making the mapping

between this operator and either the word “No” or “Not” poses a difficult problem that most children will not solve until they turn two (this despite hearing both words frequently; see Cameron-Faulkner, et al., 2007). Importantly, since children map both words to logical negation at the same time, the mapping problem appears to be the same for both. While the grammatical properties of “No” and “Not” and the speech acts each word gets used for are quite different – English “No” can be quantificational, anaphoric, a single-word prohibition or an answer to a question, while “Not” cannot – it appears that these differences do not have an effect on children’s learning to comprehend them as logical operators.

Both conceptual and linguistic limitations are attested in different cases where production precedes comprehension. For some words, like numbers and the theoretical terms alive and dead, there is reason to believe that the adult-like concepts are not initially available and that children must undergo significant conceptual change to build them before the right concept-word mapping can be made (for number: see Carey, 2009; for alive and dead: see Carey, 1985). For other cases, like color words, the mapping problem may be primarily linguistic rather than conceptual. Despite early arguments that children have difficulty identifying the conceptual domain that color words refer to (Sandhofer & Smith, 1999; Kowalski and Zimiles, 2006), recent evidence suggests that children map color words to this domain early on (and therefore have the domain available to conceptual thought), but take much longer to figure out exactly what partition of color space each word maps onto (Wagner, Dobkins & Barner, 2013).

Both conceptual and linguistic explanations for the comprehension-production gap with “No” are possible. Some suggestive support for the limiting factor being linguistic rather than conceptual comes from the fact that the comprehension of both logical “No” and “Not” tracks closely with children’s emerging production of “Not”. This connection would be unprincipled if

the limitation on the mapping was caused by a missing concept. When children hear parents using “Not”, it is used to express denial only slightly more often than “No” is, and is also frequently used to indicate rejection or the failure of an action plan (Cameron-Faulkner, et al., 2007). There would be no reason why “Not” could not be produced before its logical meaning is comprehended, just like “No” is, perhaps with a distinct meaning like rejection. It is possible that once children have mapped “Not” to a concept of logical negation, they are quickly able to do the same for “No”.

Future work should test between the conceptual and linguistic factors directly. One approach would be to use international adoption as a natural experiment (see Snedeker, Geren & Shafto, 2007; Snedeker, Geren & Shafto, 2012). International adoptees who begin to learn English as toddlers have much more sophisticated cognitive abilities than native English-learning infants when they are at the same points in language development. On the comprehension side, if children lack a concept of negation until they are about 24 months old, we might expect these older children to successfully understand both “No” and “Not” at a much earlier point in their language learning. On the other hand, the problem could be in making the mapping from the words to the logical concept, with children needing to know enough language – the meanings of content words and the syntactic instructions for combining these words together with the negative particle – in order to make that mapping. In that case we might expect the older adoptees to succeed in understanding “No” and “Not” truth-functionally only after they have reached a similar degree of language proficiency as 25-month-old native learners. On the production side, if a concept of negation is acquired around 25 months, we might expect older adoptees to start producing ‘denial’ negations – the function most closely associated with the truth-functional operator – at an earlier point in language development than native learners. If the

limitation has to do with language development, denials should emerge as children learn more words, at the same rate for adoptees and native learners.

3.6.2. The different meanings of “No”

What, then, does “No” mean for children between 15 months, when a majority already produce it, and about 24 months, when they begin to comprehend it as a truth-functional operator? One possibility goes back to suggestions made in studies of children’s production transcripts. Bloom (1970) and Pea (1980a), finding that younger children produce more ‘rejection’ and ‘nonexistence’ negations than ‘denials’, posited that these are three distinct semantic representations and suggested that the acquisition of the concept of denial is derived from the two preceding concepts. As we argued above, production data alone cannot provide sufficient evidence for this conclusion. However, when this data is taken together with evidence that the comprehension of logical, truth-functional negation, comes about a year later than the production of the negation word “No”, there is now good reason to believe that these meanings are distinct and independently acquired rather than being different usages of a single extant concept.

If they are distinct concepts, how are they related? One possibility is that they are not – if mapping “No” and “Not” to logical negation is a linguistic mapping problem, not a conceptual problem, then there is no issue of needing to build the concept of the logical operator. If, however, the limitation is that there is no logical negation concept available, it is also possible that “No” initially serves as a linguistic placeholder structure for the emergence of this concept (see Carey, 2009 for this sort of argument in the domain of number word learning).

Children who have learned that “No” means rejection may notice that some of adults’ uses of that word do not seem to denote negative affect, just as saying “There are no bears on

Mars” does not tell us much about the speaker’s feelings about bears or Mars or the one being on the other. In their speech to children, adults appear to use verbal negation of all forms predominantly for two functions – to issue prohibitions, and to express denials (Cameron-Faulkner, et al., 2007). Perhaps the disconnect between the ‘rejection’ meaning children have acquired and the broader set of meanings (some of them incomprehensible) that are used by adults serve as an invitation to form a new category. This invitation might induce children to learn the concept of logical negation in one of several possible ways. Thus the language of negation may serve as input to a Quinian bootstrapping process (see Carey, 2009), with the word “No”, mapped to ‘rejection’, serving as a placeholder on which to build a logical operator. Perhaps children would begin to notice the correspondence between the truth value of a given proposition with and without a negation word (e.g. if THIS IS A BEAR is true, THIS IS NOT A BEAR is false, and vice versa), noticing also that the rejection meaning they already have forms a special case of truth-functional negation (i.e. NOT WANT).

A different possibility is that the language helps to ‘fix’ the concept by hypothesis-testing over a space of innately available possibilities that include a logical operator that changes truth value, as in Fodor’s (1975) radical nativist account. On this account, language may also play a central role in concept fixation. Without the syntactic data about the distribution of grammatical constructions in which negation words appear and without sufficient lexical and syntactic knowledge to understand what the proposition would mean without the negation operator applied, it may not be possible for the hypothesis-testing procedure to fix the right concept in thought.

Of course, even the most basic computations involved in any mental processing done by children involve some computation-internal use of logical negation. Logical NOT-gates are

among the most basic building blocks of all computation. What, if any, is the connection between these small computational elements, which can be specified as inhibitory activation even in the intracellular mechanisms of individual cells, and the concept of negation – that representation which gets mapped to the words “No” and “Not” early in the third year of life, and which combines so freely with other concepts, like BEARS and MARS? Coming up with a detailed account for how truth-functional negation could be derived from a different, non-logical meaning is a central problem for future research on the development of negation. One possibility suggested by Bloom (1970) and supported by Pea (1980a) is that truth-functional negation is not derived from rejection, but rather from a different, intermediate meaning.

3.6.1. Nonexistence vs. Denial

One of the three categories originally used in Bloom’s taxonomy of negation production is nonexistence. As Bloom (1970) and Pea (1980a) suggest, nonexistence may be a distinct meaning that emerges after rejection and before denial, and the ability to deny might be built upon the ability to express nonexistence. We have said little about it to this point, focusing mostly on the contrast between rejection and denial, yet the hiding-game paradigms used both by us here and by Austin, et al. (2014) are readily viewed as testing the comprehension of nonexistence negation rather than denial. After all, the negative clue gives information about the ball not being in one of the containers after the hiding procedure established that it might have been there. That may seem as clear a case of nonexistence negation as ever there was.

Nevertheless there are a number of reasons to think that the hiding paradigms test for the presence of truth-functional negation, perhaps not in contrast to, but in the specific context of nonexistence. First, despite Bloom distinguishing nonexistence from denial as separate meanings, and Pea associating the latter but not the former with the truth-functional logical

operator, it is not clear that nonexistence and truth-functional negation could possibly be semantically distinct. What would it look like to specify the meaning of a nonexistence concept as something distinct from truth-functional negation? One tempting option might be to say that negation specifies a set of things that does not contain the predicate being negated, so that the utterance “No juice!” uttered upon opening the fridge means something like, 'the things in the fridge each have the property of not being juice'. As semanticists have often pointed out, this analysis of the meaning of negation quickly runs into problems (Geach, 1972; see Heim and Kratzer, 1998). Compare the sentences:

(1a) There are no bears on Mars

(1b) There is a bear named Smokey on Mars

If we treat both noun phrases in 1a and 1b as denoting sets of entities, in the case of 1b we would analyze that phrase as specifying a set containing a bear. Composing it with the predicate ON MARS, the meaning of the sentence would be: There is something that is a bear named Smokey and that thing is also on Mars. That reflects the meaning of 1b just fine. But if we run the same analysis for 1a, we get gibberish. Taking NO BEARS to mean the set of things that are not bears and combining it with the predicate ON MARS, the meaning of the sentence would be: there are things that are not bears and those things are on Mars. This is clearly not the right meaning, at least not for adults. To posit that children have a separate 'nonexistence' negation along these lines would be to posit that this incorrect meaning is exactly the one children would derive from “There are no bears on Mars”. While this has not been tested empirically, we propose that it is extremely unlikely children derive this kind of incorrect interpretation systematically.

The solution, semantically, is to identify “No” not with a function that takes some subsection of a set (that is, none of it), but as a function that relates two sets to each other. By this analysis, the “No” in “There are no bears on Mars” specifies a relation between the set of bears and the set of things that are on Mars – it specifies that these two sets do not intersect (a bit more technically, that their intersection is the empty set). With this denotation of negation, the meaning of “No juice!” in the context of opening the fridge is: the set of juice and the set of things in the fridge do not intersect. That gives the right meaning, but it is also just the standard semantic denotation for truth-functional negation. For anyone claiming that nonexistence negation and truth-functional negation are distinct, the challenge would be to define a distinct meaning for nonexistence that makes the right predictions about how it is used and understood by children.

Even if an adequate denotation for ‘nonexistence’ could be given, and if the meanings of ‘denial’ and ‘nonexistence’ negation really are separate, there is another reason to think that the search task used here and by Austin, et al. relies on understanding denials. Children begin to produce denial negations reliably between 24-30 months, by various estimates (Bloom, 1970; Pea, 1980b, 1982; Hummer, Wimmer and Antes, 1993; Choi, 1988). The ages over which children begin to succeed in these tasks match the ages over which they begin to produce denials, and not the age at which they produce ‘nonexistence’ negations, which is significantly earlier (Bloom, 1970, Pea, 1980a; Choi, 1988).

3.6.2. Denial and truth-functional negation

While here we have equated denial negation uses with the truth-functional concept, a number of researchers have suggested that children’s denials are metalinguistic rather than truth-functional (Hummer, et al., 1993; Moll, 2013; Guidetti, 2005). A metalinguistic negation is a

judgment that the negated statement is in some way conversationally inappropriate. It is indeed plausible that at least some of the denial negations produced by children are of this kind. When an adult points to a picture of a bear and calls it a cat, causing the child to say “No! Bear!” (Hummer, et al., 1993), the child may not be representing that it is false that the picture depicts a cat. Instead, the child might notice that no cats are present or had been mentioned previously, and anyway, that it is weird to call this thing that is clearly a bear anything but. Children’s “No!” may therefore be commenting on the inappropriate speech act of the adult rather than its truth value. In this way, a metalinguistic rejection is closest to ‘rejection’ in meaning – it is a labeling of an aversive response, but with the response being to a speech act rather than an offer or an order.

While it is plausible that a child saying “No!” is making a metalinguistic denial, it is harder to see the comprehension of sentence-internal negation as the comprehension of metalinguistic speech acts in the present experimental context. Perhaps in Experiment 3, where the parent asks, “Is it in the bucket?” and the experimenter answers “No, it’s not”, the negation could be interpreted as metalinguistic. But what metalinguistic interpretation could be given to the negative utterances in Experiments 1 and 2, where the experimenter simply says, “It’s not in the bucket?” To see this as a metalinguistic denial, the child would need to think that the experimenter is commenting on the inappropriateness of the claim that the ball is in the bucket; but no such claim had been made, except in the clause that is negated by the Experimenter herself. That children correctly understand the negated sentence and search in the other location suggests that they understand negation as a semantic, truth-functional operator. And in turn, that the ages at which children increasingly produce denial negations match the ages at which they

come to understand both “Not” and “No” in the search paradigm suggests that denials reflect the use of the same logical operator in production.

3.6.3. Affirmatives and the processing of negation

While the block order effects we find are less robust than the ages of success and failure, depending as they do on the exclusion criterion used, they are consistent across Experiments 1-3. That there are block order effects suggests that the 24-month-old group does not perform at chance merely because it consists of some kids who have mapped the words “No” and “Not” to the logical concept and some who have not, but that children go through a genuine transitional phase – a point at which they have mapped “No” and “Not” to the logical concept, but do not consistently deploy it when they get negative trials first. Why would getting affirmative trials first help these children’s performance?

One possibility is semantic – perhaps constructing representations of negative statements is a semantically taxing process, with a negation operator being particularly difficult to integrate into the meaning of a proposition. If constructing a negation involves first constructing an affirmative and then negating it, so that practice with the affirmative decreases the overall workload involved in computing the negation (Kaup, et al., 2006; Clark & Chase, 1972; 1974; Carpenter & Just, 1975).

Another possibility is pragmatic – perhaps negatives are more plausible or natural within a discourse following a series of affirmatives, and similar affirmatives in particular (see Flusberg & DeVilliers, 1975; Wason, 1965). Indeed, proponents of this account have argued that many studies pointing to the general difficulty of comprehending negation do not provide a sufficiently

supportive discourse context. Hearing a series of similar affirmative assertions lends pragmatic support and contextual plausibility to a relevant negative assertion appearing later on in the discourse, and this appears to facilitate the processing of negation for adults (Niewland & Kuperberg, 2008; Tian Breheny, & Ferguson, 201).

Testing between these two possibilities would require isolating semantic and pragmatic factors. One approach would be to make the discourse context as supportive as possible for the introduction of a negation, without actually introducing the affirmative proposition into the discourse. If this eases comprehension success, it would point to the involvement of pragmatic factors. If it eliminates the difficulty with processing negation entirely, it would point to the lack of purely semantic difficulty.

3.6.4. What underlies the representation of emptiness?

When 20-month-olds who fail in comprehending the truth-functional meaning of “No” and “Not” successfully avoid a bucket they have seen is empty, what underlies their success? What, if any, is the relationship between the representation generated when seeing the empty bucket and the representation only older children build when hearing that the ball is “not in the bucket?” One possibility is that 20-month-olds in Experiment 4 succeed on negative trials by using a logical negation operator, representing that “the ball is not in the bucket”. In this case, their failure given verbal clues would indicate that they have not yet mapped the words “No” and “Not” to the logical operator concept that they possess. Alternatively, their success with visual cues could be due to a non-logical representation – perhaps a representation of emptiness, or a representation in a non-propositional format, such as visual imagery, which may not be able to support the use of a logical negation operator.

One way to choose between these possibilities would be with a non-verbal test of logical negation. Several studies that have presented pre-linguistic with tasks where they could succeed by deploying a negation operator have found a failure to do so. Feiman, Carey & Cushman (2015) found that although 7- and 14-month-olds represent APPROACH, but fail to represent AVOID, which they could compose as NOT-APPROACH if they possessed a negation operator able to participate in combinatorial thought. Similarly, Hochmann, Mody & Carey (under review) found that 14-month-olds represent SAME but not different, which they could have composed as NOT-SAME using a negation operator. Another clue is that children also begin to succeed on some tasks that may involve a negation operator around 18 months, such as using the mutual exclusivity criterion for word learning (Halberda, 2003), close to the age at which they succeed in the search task of Experiment 4. Taken together, these results point to the possibility that the representation of an empty container in Experiment 4 could deploy a logical negation operator, perhaps in a representation like THE BALL IS NOT IN THE BUCKET. However, a more direct test of the existence of a domain-general logical negation operator pre-linguistically is urgently needed.

Chapter 4:

The logic in language: How all quantifiers are alike, but each quantifier is different

4.1. Introduction

All languages have systematic rules of interpretation, allowing listeners to derive complex sentence meanings from combinations of words. The simplest hypothesis is that the rules that govern semantic and syntactic composition are perfectly coupled, so that every different meaning is expressed by a unique natural language sentence (Montague, 1968). Semantic ambiguities – cases where the same sentence has two different meanings – challenge this simple hypothesis, and thus are central to our understanding of the relationship between meaning and form. A particularly systematic type of ambiguity, known as a scope ambiguity, has been studied extensively by semanticists. A scope ambiguity arises whenever two quantifiers (like EVERY and A) occur in the same clause. For example, take the sentence:

(1) Every kid climbed a tree

This sentence could mean either that there is a single tree, climbed by every kid, or that every kid climbed some tree, but no two kids necessarily climbed the same one. The ambiguity in this sentence does not reflect an ambiguity in the speaker's mind, since we surely know which meaning we want to convey. Thus we must have some format of representation that is less ambiguous than the English sentence and which precedes language production. In many semantic theories (see Hornstein, 1984; May, 1985; Heim & Kratzer, 1998 inter alia), this is captured by positing a level of representation, Logical Form (LF), separate from the surface form of the sentence, where the two interpretations are distinct. The two LFs for (1), ignoring tense, are presented below, with corresponding paraphrases:

(2a) $\forall x[\text{Kid}(x) \rightarrow \exists y[\text{Tree}(y) \wedge \text{climbed}(x,y)]]$

For every x, if x is a kid, then there exists a y, such that y is a tree and x climbed y

(2b) $\exists y[\text{Tree}(y) \wedge \forall x[\text{Kid}(x) \rightarrow \text{climbed}(x,y)]]$

There exists a y, such that y is a tree, and for all x, if x is a kid, then x climbed y

These LF representations are not ambiguous the way the English sentence (1) is because they specify the order in which the two quantifiers bind their variables. (2a) has the universal quantifier EVERY taking the widest possible scope, with the value for y being fixed relative to each choice of the value for x. This is the interpretation with potentially many trees – for every kid, a different tree. We refer to this as a “Universal-wide” or “U-wide” interpretation. (2b) is the interpretation where there can be only one tree. The variable for the tree, y, is fixed first, and then the universal quantifier ranges over many values of x in relation to that y. This sort of LF has the Existential quantifier, A, taking wide scope, and we refer to it as an “Existential-wide” or “E-wide” interpretation. Sentences like (1) are thus scopally ambiguous because the two possible meanings arise from the relative scope of the two quantifiers EVERY and A in LF.

The factors that affect how scope ambiguities are interpreted have been a topic of extensive study and debate within linguistics, but only recently has the application of psycholinguistic methods been used to ask about the content of LF representations (Raffray & Pickering, 2010; see also, Chemla & Bott, 2015). Raffray and Pickering were interested in how the construction of one LF representation affects the construction of another and what these patterns can tell us about the information that is relevant for constructing a logical form. Their study focused solely on sentences with EVERY and A. Participants read scopally ambiguous sentences like (1), and then picked, from two pictures, the one they thought best matched the sentence. On the prime trials, participants were forced to pick a picture corresponding to one scopal interpretation (e.g. the U-wide interpretation) because the other picture mismatched the sentence based either on the subject or object noun. On the target trials, which immediately followed the primes, participants read a new scopally ambiguous sentence, like (3) with the same two quantifiers, but different nouns.

(3) Every hiker climbed a hill

Participants were then given a choice between two pictures, one for each of the scopal readings of the target sentence (see Fig. 4.3). Raffray and Pickering found that participants were more likely to pick a U-wide target picture after a U-wide prime than an E-wide prime. They interpret this priming effect as evidence of shared representational resources across the LFs constructed from the prime and target sentences.

Our experiments explore the locus of this priming effect, asking whether there are shared mental structures across sentences with different quantifiers and those with different verbs. We have two main goals. One is to home in on the nature of the operations used to construct different scopal interpretations, exploring what is represented explicitly by the combinatorial semantic machinery, and what kinds of content are irrelevant for these operations (what distinctions they abstract away from). For example, is it the case that the specific lexical entries of verbs are called upon as part of the combinatorial process that assembles quantifiers in different scopal relations to each other, or are individual verb meanings irrelevant to the scopal operations that are being primed in this procedure? The other, complimentary goal is to use our findings about the degree of abstraction in LF to speak to longstanding assumptions in the formal semantics literature about the distinction between conceptual content and the combinatorial properties of meaning, as well as to debates on whether different quantifiers with similar meanings (e.g., universals) have different combinatorial properties and mechanisms for assigning scope. Specifically we will explore: whether some quantifiers necessarily force a U-wide or E-wide reading (and therefore do not, in fact, give rise to scope ambiguities), whether some types of readings are systematically more difficult to construct than others, and whether particular

quantifiers (EVERY, ALL, and EACH; TWO, THREE and FOUR) share a common mechanism of scope assignment.

In the remainder of the introduction, we provide a brief outline of linguistic theories of quantifier scope, discuss how Raffray and Pickering's experiments bear on these theories, and outline our plan to exploit this paradigm to answer new questions.

4.1.1 Theoretical approaches to quantifier scope

How listeners derive an unambiguous representation from an ambiguous surface structure like (1) is a matter of longstanding debate (for a review see Ruys and Winter, 2010). Several classes of theories have been proposed. All posit additional operations on top of syntactic parsing of the surface structure, but they differ as to whether these operations are syntactic (e.g. Quantifier Raising or QR, Chomsky, 1976; May, 1977; 1985; Quantifying-in, Montague, 1972; Rodman, 1976) or semantic (e.g. Cooper storage, Cooper, 1983; Keller, 1988; type-shifting of expressions, Hendriks, 1988; type-shifting of composition rules, Barker, 2002) or involve a more complex mapping at the syntax-semantics interface (e.g., a lexicalized grammar where different quantifier words have different scopal mechanisms, Beghelli and Stowell, 1997; Steedman, 2012; See also Reinhart, 1997; Winter, 2001).

Evidence for and against different theories of quantifier scope has come from patterns of judgments about the availability of different readings for scopally ambiguous sentences. For example, *prima facie* evidence of parallel constraints on the movement of both Wh-phrases and quantifiers out of relative clauses provided some of the early evidence in favor of quantifier scope relying on syntactic operations like QR, but no theory yet exists that can easily account for all of the intricate patterns of preference and grammaticality judgments of different readings (see

Ruys and Winter, 2010). Indeed, finding such a theory has been one of the key projects of semantics.

It is worth noting that the term “Logical Form” is specific to the syntactic theories, though we will be using it in a theory-neutral manner. Following Hornstein (1984) and May (1985), LF is commonly described as a level of representation that is separate from surface syntax, directly interpreted by a semantic mechanism, but not directly connected to phonetic form (with operations like QR therefore being covert, as opposed to overt syntactic operations like Wh- movement). Our use of the term LF throughout this paper does not indicate a commitment to this framework. All theories require some representations that disambiguate between the kinds of readings given by (2a) and (2b). We refer to these unambiguous representations as LF, and to the derivation of unambiguous meaning as the construction of LF, both for convenience and because the LF tradition has been the most dominant one in discussions of scope ambiguity. Importantly, however, the questions that we address in this paper arise in all theories of scope ambiguity, even though the terminology changes.

4.1.2 Accounting for scope preferences

Prior linguistic work has identified three factors that may influence which reading of a scope ambiguity is preferred. First, the order of the quantifiers in the sentence has been argued to be a major factor in their relative scope assignment, with quantifiers earlier in the sentence preferring wide scope (Johnson-Laird, 1969; Lakoff, 1971). In languages like English, the earlier quantifier is typically the quantifier that is higher in the syntactic tree. When these factors are disentangled, some theorists have argued that it is the higher quantifier (not the earlier one) that takes wide scope (Jackendoff, 1972; VanLehn, 1978; Lidz & Musolino, 2002). This observation is broadly consistent with syntactic theories (e.g. May, 1985), which posit that a costly raising

operation is required when a quantifier that is lower in the syntactic tree (like an object quantifier) takes wide scope. If we assume that the parser avoids costly operations, this predicts a preference for linear scope readings (see Lidz & Musolino, 2002 for discussion).

Second, other theorists have suggested that scope preferences have a lexical component. In her seminal work, Ioup (1975) observed that different quantifier words seem to fall along a hierarchy based on the scope they prefer to take (see also VanLehn, 1978; Vendler, 1967; Quine, 1960 for related observations). This hierarchy is reproduced in (4) with quantifiers preferring wider scope appearing on the left:

(4) EACH > EVERY > ALL > MOST > MANY > SEVERAL > SOME > A FEW

Ioup observes that, in general, quantifiers that select a larger set size seem to prefer wide scope over those that select smaller sets, but she makes no claims about why that should be. Since then, nearly every study of the comprehension and production of scopally ambiguous sentences has found lexically-based differences in the preferred scope assignments of quantifiers—though most of these studies have only examined sentences containing EVERY and A. EVERY prefers to take wide scope over A (generating the U-wide reading) regardless of the linear or hierarchical order of the quantifiers (Chemla & Bott, 2015; Clark & Kar, 2011; Raffray & Pickering, 2010; Bott & Radó, 2007; Filik, Paterson & Liversedge, 2004; Gillen, 1991; Micham, Catlin, VanDerven & Loveland, 1980; Catlin & Micham, 1975). The one exception to this generalization is a study by Kurtzman and MacDonald (1993), which found a slight preference for A to take wide scope over EVERY. This finding, however, has been argued to reflect factors specific to the dependent measure that they used (see Tunstall, 1998, for discussion). Thus the broader generalization appears to be secure.

However, it is not clear that the observed differences between sentences containing different quantifiers are caused by stable lexical properties that play a role in scope assignment. The third and final possibility is that scopal disambiguation is based on more general conceptual knowledge or expectations about the sorts of things that speakers are likely to say (Katz, 1980; Fodor, 1982; Van Berkum, Brown, & Hagoort, 1999; Altmann & Steedman, 1988). For example, if the context suggests that there might be a single referent for a singular indefinite (like A TREE) then the listener is more likely to get an E-wide reading. But if the context suggests that multiple referents would be needed, then a U-wide reading would be preferred. On this hypothesis, differences in the number of things being discussed affect the plausibility of the message and so different quantifiers will show different patterns of interpretation, even though lexical knowledge plays no direct role in comprehension. For example, the observation that EVERY often takes scope over A could merely be a side effect of the fact that the discourses we encounter and construct with these words are more apt to support a U-wide interpretation than an E-wide one.

4.1.3 The abstractness of the operation

The present paper uses priming to explore the nature of the operations involved in disambiguating scope, following Raffray and Pickering (2010). These questions are separate from the debates between QR, type-shifting, and other linguistic theories, although they bear on those issues. Whatever theory one takes, there is a question about the abstractness of the disambiguating operations. For example, if the E-wide LF in (2b) is constructed by raising the object quantifier to a higher scope position in LF, then there is a question of how many raising operations there are and how they work – is there a single operation that raises the object quantifier or the second quantifier in a sentence, whatever it happens to be? Or is there a single

operation for raising universal quantifiers but a separate operation for raising number quantifiers? Or are there different raising operations for each quantifier word? These same questions arise for a type-shifting account, only phrased in terms of the nature of the type-shifting operations. And perhaps some quantifiers are subject to movement by QR while others take wide scope through type-shifting or another mechanism.

The factors that influence preferred scope readings bear on the issue of the abstractness of the scoping operations. Ioup's observations give us reason to suspect that there might be different operations for different quantifiers – the operation that acts on EACH and preferentially gives it wide scope might be different than the operation acting on A, which prefers to assign it narrow scope. That evidence is only suggestive, however. It could also be the case that there is a single abstract operation, but that quantifier words vary in the degree to which they call upon it. And as we mentioned above, it is possible that these differences reflect expectations about different situations in the world or the things others are likely to talk about, with different quantifier words being correlated with these expectations.

Raffray and Pickering's findings (2010) provide some initial insights into properties of the operations involved in constructing LF. Their first experiment, described above, demonstrates that scopal relations can be primed when the quantifiers are held constant but the nouns vary. In the subsequent experiments they established that the priming effect holds when the prime sentence is passivized, while the target stays active. The U-wide interpretation of a sentence like, "*A tree was climbed by every kid*" primes the U-wide interpretation of a sentence like "Every hiker climbed a hill". This pattern demonstrates that the primed operation (or representation) is sensitive to the underlying agent (the deep subject) but unaffected by changing the syntactic position of the arguments. These results, however, leave open the question of

whether LF construction involves a number of different operations that are specific to individual quantifier words, like EVERY and A, or is one general operation that applies to many or even all quantifiers.

One way we address this question in the present work is by asking what role lexical content plays in LF construction. Are lexical items merely one cue (among many) that bias us toward particular scope relations? Or is LF construction differentiated according to the lexical or semantic content of quantifiers? This investigation also provides an opportunity for looking at the meaning shared between different quantifiers – if an LF is abstract relative to the universal quantifiers, but not others, it would suggest that the universals have an underlying representational similarity that plays a role in scope assignment and LF construction.

In Experiment 1, we begin with baseline measures of how scopally ambiguous sentences with different quantifiers are interpreted, when not preceded by any prime. We test the three English universals EVERY, EACH, and ALL, as well as numbers like THREE, FOUR and FIVE. This will allow us to test whether these different quantifiers show different scopal preferences, as suggested by Ioup (1975) and others. It will also provide a baseline for subsequent priming effects. In Experiment 2, we take the same quantifiers and use them as both primes and targets in order to look at whether the priming effect found by Raffray and Pickering extends from one universal quantifier to another (e.g. EACH to ALL, EVERY to EACH, etc.), holding constant other factors like discourse, noun and verb content. We also examine whether there is a priming effect at LF between quantifiers from different families – from a universal in the prime to a number in the target, and vice versa. We find that priming seems to be limited to cases where the same quantifier appears in the prime and the target. Experiment 3 rules out the hypothesis that this pattern simply reflects a sudden drop off in priming due to the number of words that the prime

and target have in common. We compare prime and target sentences that share the same quantifiers and have either the same verb or different verbs, and find robust priming effects in both cases. Experiment 4 looks at priming across sentences with different numbers. Although they are different lexical items, they are highly semantically related. We find priming that extends from one number to another.

4.2. Experiment 1: Baselines

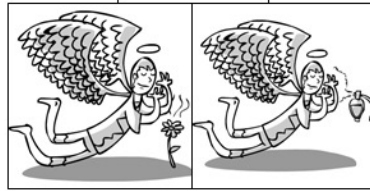
4.2.1 Method

4.2.1.1. Participants

For every condition in every experiment reported below, we recruited 128 unique participants on Amazon Mechanical Turk. Participants had to have had 98% of their previous work approved and had to be logging in from an IP address within the United States. Participants were barred from participating in more than one condition across all the experiments reported here. Participants were excluded for answering less than 90% of filler trials correctly in all experiments or if they indicated that English was not their first language. After these exclusions were applied, there were 116 participants in each of the EVERY (Mean age: 30.7, Range: 18-78), ALL (Mean age: 30.4, Range: 18-66) and number (Mean age: 34.8, Range: 19-73) conditions, and 119 in the EACH condition (Mean age: 30.4, Range: 18-57), who were included in the results below.

Filler Trials: Sentences and Pictures

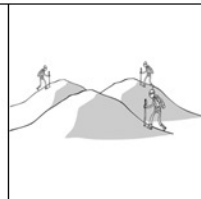
Filler Trial



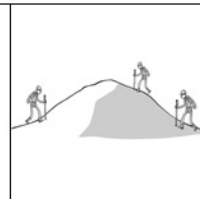
Filler and prime replacement for Experiment 1:
The angel smelled the flower

Target Trials: Sentences and Pictures

Universal-wide Target Choice



Existential-wide Target Choice



Experiment 1: *Every/Each/All of the/Three hiker(s) climbed a hill*

Fig. 4.1.: Sample stimuli for Experiment 1. Filler trials (top) always directly preceded target trials (bottom). The correct picture choice on the filler trial depicted the corresponding reading of the sentence while the incorrect choice pictured a noun not present in the sentence.

4.2.1.2. Stimuli

Experiment 1 used 24 target items, taken from Raffray and Pickering (2010, R&P hereafter), each of which consisted of a scopally ambiguous sentence (e.g. “*Every hiker climbed a hill*”) and two pictures. One picture depicted a U-wide reading (e.g. three hikers, each climbing a different hill), and the other depicted the E-wide reading (e.g. three hikers climbing the same hill). An example of a target trial is shown in Figure 4.1. Depending on the condition,

the subject quantifier in the sentence was either EVERY, EACH, ALL, or one of the numbers (THREE, FOUR or FIVE). The object quantifier was always A.

As in R&P's Experiment 1, all of the sentences contained 12 past tense verbs (e.g., CLIMBED, CHASED, WATCHED) in an active construction, with animate subject nouns and either animate or inanimate objects in equal halves (E.g. "*Three hikers climbed a hill*" or "*Four sailors saw an airplane*"). There were either three, four or five subject entities (with a third of the trials having each number. The number of object entities was either one (for the E-wide pictures), or the same as the number of subject entities (for the U-wide pictures; see Fig. 4.1). The side on which the U-wide and E-wide pictures appeared was counterbalanced within subjects.

Target trials were separated by three to five filler trials. There were 96 unambiguous filler trials in total. These, like target trials, consisted of a sentence and two pictures. However, fillers had unambiguous transitive sentences (e.g. "*The angel smelled the flower*"; see Fig. 4.1) which matched one of the pictures, while the other picture depicted a different subject or object. Of the fillers, 72 items were taken from R&P, altered only to make all sentences transitive, so that they would more closely resemble target items. The remaining 24 fillers were new items, constructed by reusing some of the picture pairs from R&P's fillers, but changing the sentence so that the picture choice that was previously a distractor would be correct (e.g. "*The angel smelled the perfume*" for the same sample filler pictures in Fig. 4.1). To familiarize participants to the forced-choice task, an additional seven practice trials were provided. These had the same structure as the fillers. Stimuli were presented in one of two orders, one the reverse of the other.

4.2.1.3. Procedure

Participants first filled out a demographic questionnaire that asked, among other questions, whether English was their first language. They then completed the seven practice

trials before proceeding with the experiment. On each trial, participants read a printed sentence centered above two pictures. Participants were instructed to select the picture that matched their interpretation of the sentence. They made their selection by clicking on one of the radio buttons located underneath each picture. Participants could not click on both buttons, and could not advance to the next trial until they clicked a button. Each trial was presented on a separate page, so that participants could not see multiple trials at once or return to a previous trial.

4.2.2 Results and Discussion

Our dependent variable was whether the participant selected the universal wide picture, which we will call a U-wide response, on the target trials. Results for each quantifier condition are presented in Figure 4.2. The data was analyzed in the R programming language, v3.1.0 using the lme4 package, v1.1-8 (Bates, et al., 2015) to build a logit mixed-effects model (see Jaeger, 2008) with the maximal random effects structure appropriate for this experimental design (Barr, et al., 2013). Table 4.1 shows the results of the statistical analyses and Figure 4.2 shows the U-wide response rates by condition.

Target Quantifier, in this experiment, was the fixed variable of interest, with four levels corresponding to the different quantifier words tested. In our initial omnibus model we treated the ALL condition as the baseline and included three predictors, one for each of the remaining conditions (i.e., a dummy coding scheme). In addition, the model included subject and item as random effects with random intercepts. In this analysis, EACH and EVERY resulted in more U-wide responding than ALL, while the numbers results in less U-wide responding than ALL (see Table 4.1). Follow-up analyses contrasted pairs of conditions in descending order of U-wide response rate: EACH vs. EVERY ($z=3.29$; $p=0.001$); EVERY vs. ALL ($z=15.1$, $p<0.0001$); ALL vs. numbers ($z=-9.55$; $p<0.0001$).

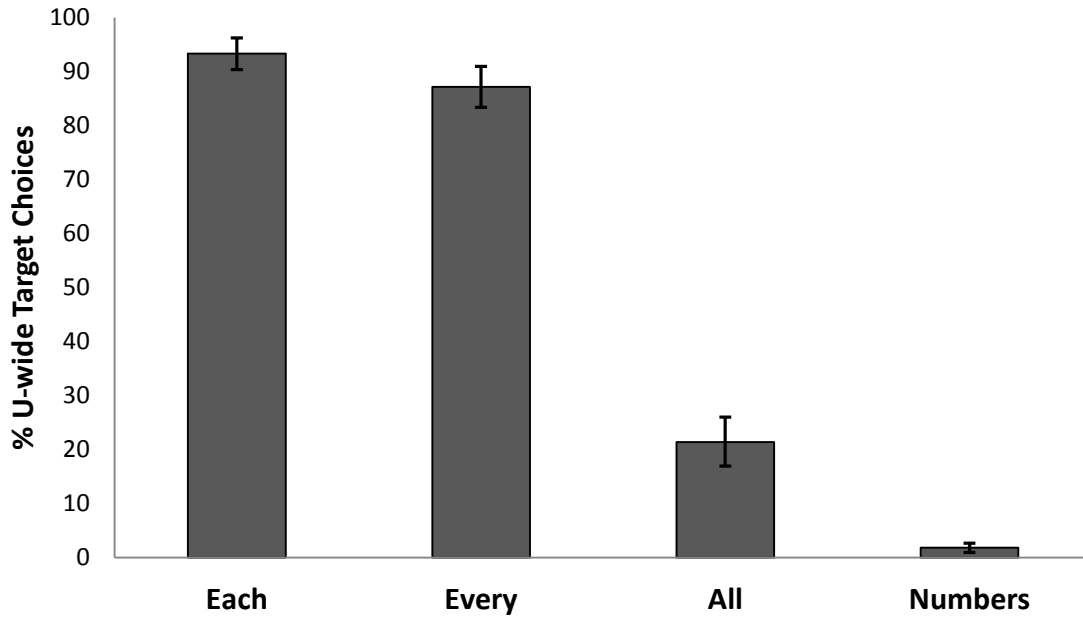


Fig. 4.2.: Participants' rates of U-wide choices on ambiguous target trials in Experiment 1. Error bars indicate 95% confidence intervals, averaged across items with subjects as the random variable.

Predictor	Coefficient	Wald Z	p-value
Intercept	-2.62	-7.35	0.0001 ***
Condition = Every Baseline	6.64	14.82	<0.0001 ***
Condition = Each Baseline	8.14	16.81	<0.0001 ***
Condition = Number Baseline	-4.31	-9.58	<0.0001 ***

Table 4.1: Logit mixed effects omnibus model for Experiment 1, with all quantifier baselines and the ALL condition as the dummy-coded contrast.

These findings strongly support earlier claims, grounded in linguistic intuitions, that different quantifier words strongly influence scope resolution (Ioup, 1975; Vendler, 1967; Quine, 1960). At first glance these results appear to confirm Ioup's hierarchy, in which EACH has the strongest wide-scope preference, followed closely by EVERY, with ALL showing a preference for a narrow scope reading (i.e., an E-wide interpretation). But critically, Ioup's proposal is about the relative scope preferences between these different quantifiers (e.g., that EVERY will take scope over ALL). Our findings demonstrate, instead, that there are differences in how likely these quantifiers are to take scope over another quantifier, the indefinite A.

Ioup observes that the ordering of quantifiers in her hierarchy can be characterized as a tendency for quantifiers picking out large sets to take scope over those picking out smaller sets. Ioup therefore predicts that universal quantifiers will consistently take scope over singular indefinites. This prediction is correct for EVERY and EACH but not for ALL, which was often given an E-wide interpretation, even though ALL picks out sets that are just as large, if not larger (see Tunstall, 1998) than the others. Similarly, on Ioup's proposal, numbers should take scope over A (since THREE clearly picks out more than one entity) but we instead find that the numbers had an extremely strong E-wide bias. In the General Discussion, we return to the implications of these findings for accounts that emphasize the influence of structural rather than lexical factors on scope preference.

We now proceed to the heart of the investigation, exploring whether and to what degree different quantifiers share common mechanisms of scope assignment by looking at patterns of priming. The results of Experiment 1 will provide a baseline for evaluating the priming effects of Experiment 2.

4.3. Experiment 2: Between and within quantifiers

4.3.1. Method

4.3.1.1. Participants

An additional 128 unique participants were recruited for every prime-target quantifier pairing. The exclusion criteria were the same as for Experiment 1. After these exclusions were applied, the means of ages ranged between 29.4 and 33.1 across conditions, with comparably wide age ranges to Experiment 1. Table 4.2 shows the numbers of participants in each prime-target quantifier pairing after the application of the same exclusion criteria as in Experiment 1.

	Every	Each	All	Number
Every	112	112	108	116
Each	118	123	121	120
All	110	120	111	110
Number	108	118	117	110

Table 4.2: Numbers of participants in each prime-target quantifier pairing, after exclusion criteria were applied. Columns indicate the quantifier word in the target sentences and rows indicate the quantifier in the prime.

4.3.1.2. Stimuli and Procedure

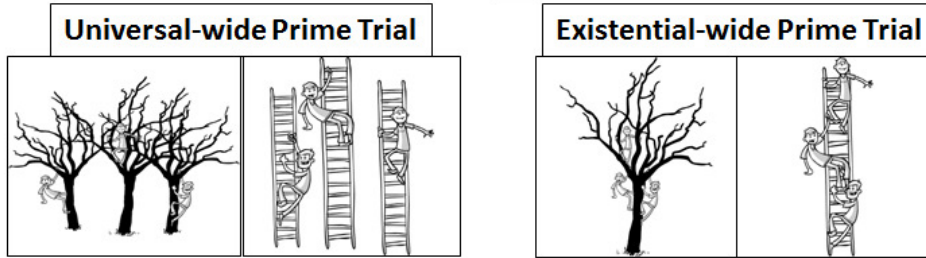
The stimuli of Experiment 2 were similar to Experiment 1. We used the same target trials, preceded by novel primes. The prime trials took the place of the extra fillers created for Experiment 1, and were placed immediately before the corresponding target trials. The target trials and the other fillers were unchanged from Experiment 1. Each prime sentence was accompanied by a prime picture that matched the sentence (but forced the reading that was being primed) and a distractor picture, which mismatched the sentence on the subject noun (50%) or the object noun (50%), but matched the correct picture in terms of its quantifier-scope assignments (see for example, the U-wide and E-wide distractor items in Figure 4.3). The stimuli (sentences and pictures for both primes and targets) were taken from R&P, modifying only the quantifier words in the primes and targets.

Across conditions, we manipulated the quantifier words used in the subject of the prime and target sentences (EVERY, EACH, ALL,¹⁸ or the numbers THREE, FOUR or FIVE) for a fully crossed 4(Prime Quantifier) X 4(Target Quantifier) between-participants design. Within participants, we manipulated Prime Scope (U-wide or E-wide).

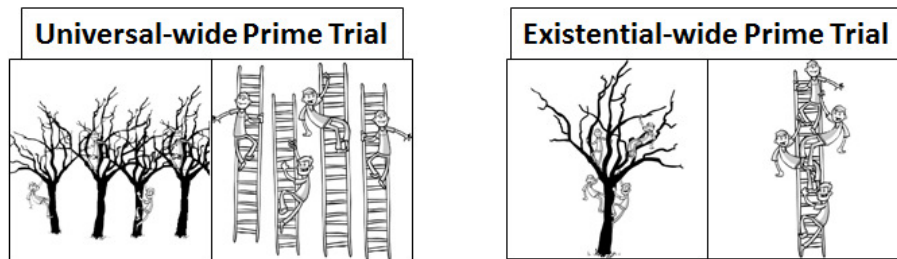
¹⁸ We used the partitive ALL OF THE instead of ALL to ensure that participants interpreted the sentence as referring solely to the depicted referents. “*All kids climbed a tree*” could be interpreted as a universal generalization about children as a kind (e.g., All dogs descended from wolves), and under this interpretation the E-wide interpretation is wildly implausible.

We counterbalanced (within subjects and across items): the side of the correct prime picture, the side of the matching target picture, and whether the correct prime picture and matching target picture were on the same or opposite sides. The Prime Scope (U-wide or E-wide) was manipulated within item and within participant by creating two counterbalanced lists. Two orders were generated as in Experiment 1 with the constraint that no more than two trials in a row had the same Prime Scope. In all other respects, the procedure was identical to Experiment 1.

Prime Trials: Sentences and Pictures

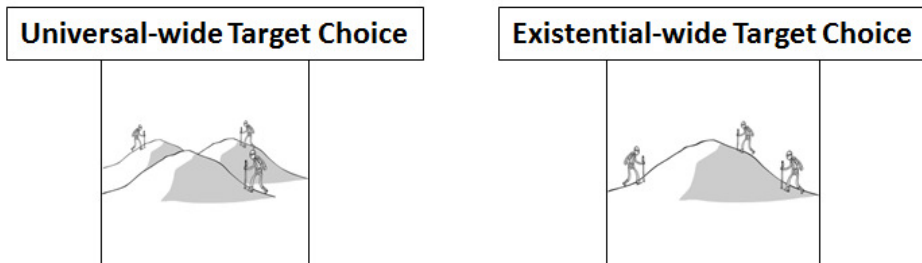


Experiment 2: *Every/Each/All of the/Three kid(s) climbed a tree*



Experiment 4: *Four kids climbed a tree*

Target Trials: Sentences and Pictures



Experiments 2 and 4: *Every/Each/All of the/Three hiker(s) climbed a hill*

Fig. 4.3: Sample prime and target items for Experiments 2 and 4. Prime trials (top) always directly preceded target trials (bottom). There were two types of prime trials – U-wide and E-wide. The correct picture choice on the prime trial depicted the corresponding reading of the sentence while the incorrect choice pictured a noun not present in the sentence. Participants in the within-quantifier conditions of Experiment 2 saw prime and target sentences with the same quantifier words while participants in the between-quantifier conditions saw a different quantifier word in the prime than the target sentence. In Experiment 4, prime pictures were modified to show a different number of subject entities (e.g. extra kids) so that the prime sentence used a different bare numeral than the target (e.g. “Four kids climbed a tree” instead of three), while the target trial remained the same (e.g. “Three hikers climbed a hill”).

4.3.2. Results and Discussion

As in Experiment 1, the data was analyzed using logit mixed effects models. For all of the analyses described below, we modeled response type on the target trials (U-wide or E-wide), with random intercepts and random slopes relative to prime scope for subjects and items (uncorrelated with each other).¹⁹

Our central question is how different primes might influence the processing of the target sentences that follow them. From Experiment 1, we know that there are large baseline differences in preferred reading between target sentences with different quantifiers. Therefore, to look for priming effects, we hold the target sentence constant and look for changes in participants' responses as a factor of both the type of prime preceding any given target (U-wide or E-wide) and the different quantifier words in the prime sentences. Critically, this analysis strategy allows us to focus on responses to the exact same target sentence and explore whether they vary based on the preceding primes.²⁰

4.3.2.1. Within-Quantifier Priming

First, we want to establish whether the picture priming paradigm used here can indeed affect participants' target choice when the quantifiers in the prime and target sentences match. The use of EVERY in both prime and target replicates the design of Raffray and Pickering's Experiment 1 (and see Chemla & Bott, 2015), while the other conditions extend this within-

¹⁹ In almost all of the analyses in all experiments, the inclusion of a parameter of correlation between the random slope and intercept in the model (the default setting in lme4 version 1.1-7) resulted in overparameterization of the model. This is unsurprising given that Prime Scope – the variable by which the slope was being estimated – is binary. In no analysis did the inclusion of the correlation estimate improve model fit on an ANOVA test between models. This parameter was therefore excluded from all analyses.

²⁰ We tested all of the models reported here and in Experiment 3 and 4 with and without the two-level fixed effects of Order and List. We found some significant main effects and interactions of Prime Type and Prime Quantifier with these variables, reflecting the effects of specific items and orders. In no case did the inclusion or exclusion of these effects change the significance of the effects we were interested in. For simplicity and consistency, we therefore report the analyses without these fixed effects throughout.

quantifier priming design to the other quantifiers – EACH, ALL, and the numbers THREE, FOUR, and FIVE.

We find significant priming effects Within-Quantifier for EVERY ($\beta=0.63$; $z=3.8$; $p=0.0001$), EACH ($\beta=0.74$; $z=4.89$; $p<0.0001$), ALL ($\beta=0.57$; $z=4.7$; $p<0.0001$), and a marginally significant effect for the numbers ($\beta=0.37$; $z=1.93$; $p=0.05$). In a model including all four Within-Quantifier conditions, there were no significant interactions between Prime Scope and Condition, indicating that the size of the priming effects in each of the Within-Quantifier cases did not differ significantly from each other.

4.3.2.2. Comparison of Between- vs. Within-Quantifier priming effects

Next, we want to ask whether there are also priming effects between different quantifiers – when the prime contains one quantifier and the target another – and whether these differ from the priming effects Within-Quantifier. We analyze each target quantifier separately, including fixed effects of Prime Scope (U- or E-wide), and Quantifier Overlap (Within- or Between-Quantifier) and their interaction (Table 4.3). The interactions between Prime Scope and Quantifier Overlap are significant in all cases except for the numbers.

Since we already know that there is Within-Quantifier priming for all four quantifiers in target position, we next examined only the Between-Quantifier conditions (without the Within-Quantifier conditions for each target quantifier). These analyses included the fixed effects of Prime Scope (U- or E-wide) and Prime Quantifier (with three different prime quantifiers for every target quantifier, varying by target; for example, if EVERY was in the target, the three prime quantifiers included in this Between-Quantifier analysis would be EACH, ALL, and numbers).

	Intercept	Prime Scope	Quantifier Overlap	Quantifier Overlap X Prime Scope
Every	$\beta=2.21$ $z=10.12$ $p<0.0001$ ***	$\beta=0.04$ $z=0.48$ $p=0.63$	$\beta=-0.72$ $z=-2.67$ $p=0.008$ **	$\beta=0.52$ $z=3.68$ $p=0.0002$ ***
Each	$\beta=4.27$ $z=19.59$ $p<0.0001$ ***	$\beta=0.21$ $z=1.93$ $p=0.05$ †	$\beta=-1.55$ $z=-5.81$ $p<0.0001$ ***	$\beta=0.49$ $z=2.89$ $p=0.004$ **
All	$\beta=-2.21$ $z=-8.25$ $p<0.0001$ ***	$\beta=0.20$ $z=2.84$ $p=0.004$ **	$\beta=-0.2$ $z=-0.95$ $p=0.34$	$\beta=0.34$ $z=2.4$ $p=0.01$ *
Number	$\beta=-4.62$ $z=-17.39$ $p<0.0001$ ***	$\beta=0.01$ $z=0.06$ $p=0.95$	$\beta=0.36$ $z=1.31$ $p=0.19$	$\beta=0.32$ $z=1.47$ $p=0.14$

Table 4.3: Within- vs Between-Quantifier analyses in Experiment 2. U-wide responses are coded as 1, and E-wide as 0. Each row is a separate logit mixed effects model, individuated by Target Quantifier. Each model includes a within-participant fixed effect of Prime Scope, a between-participants fixed effect of Quantifier Overlap, and an interaction between the two.

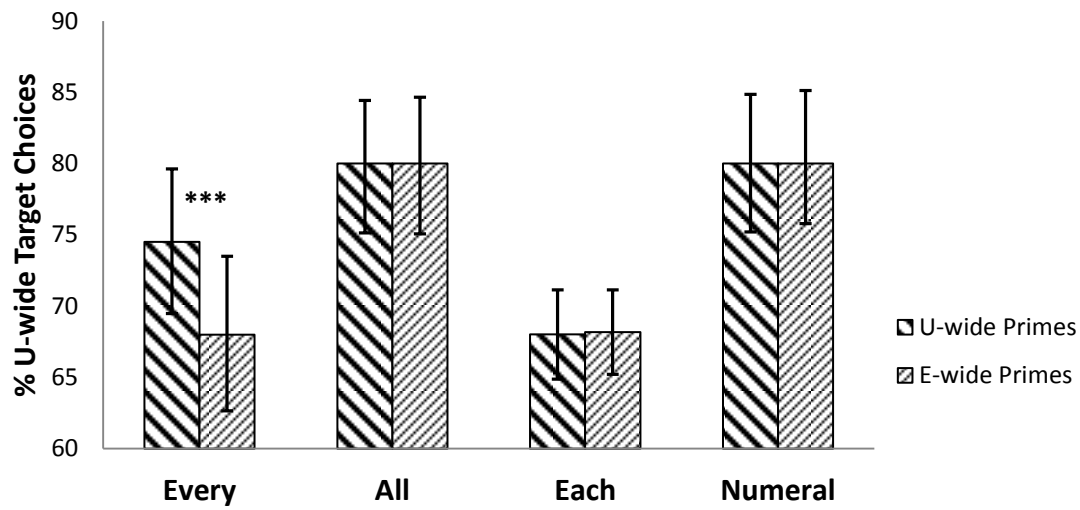
We found a significant interaction between Prime Scope and Prime Quantifier in the case of number targets. Logistic mixed effects models of each Between-Quantifier condition (ie. with different Prime Quantifiers) separately showed that this interaction was driven by a significant effect of Prime Scope in the condition where the target sentences contained ALL and the prime sentences contained numbers ($z=2.01$, $p=0.04$). Since this was a small effect and was not predicted by any theoretical account, we replicated this condition with an additional 128 participants. Including the data from both the original run and the replication, the effect was no longer significant in this condition ($z=1.51$; $p=0.13$), nor was there any longer a significant Prime Scope by Prime Quantifier interaction in the larger Between-Quantifier model that included all of the Between-Quantifier conditions with number targets. We therefore conclude that the

original effect was a false positive. We conclude that there was no robust effect of Between-Quantifier priming across any two quantifiers.

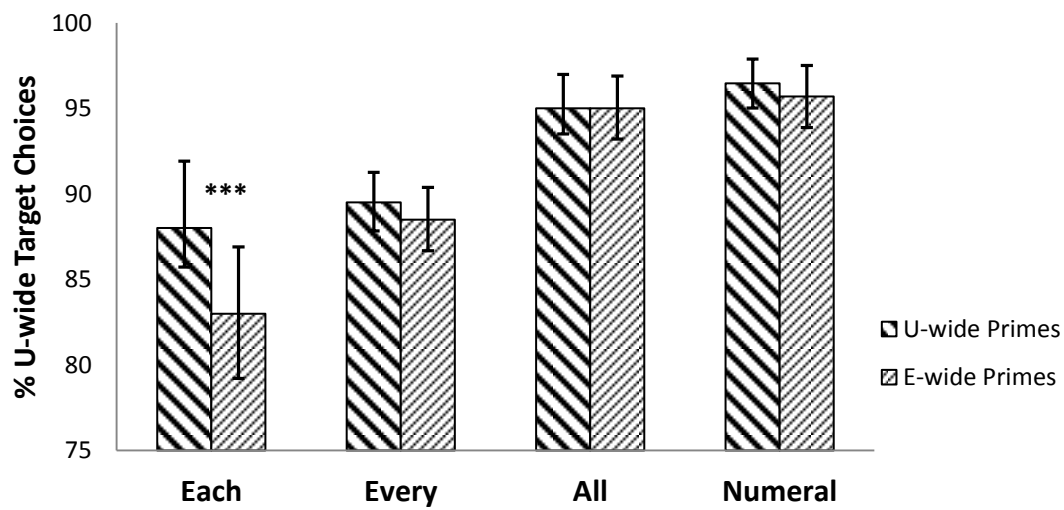
Despite the lack of main effects of Prime Scope in the Between-Quantifier analyses for different target quantifiers, another planned comparison of theoretical interest is between the different types of Between-Quantifier cases. For example, if the quantifier in the target sentence is EVERY, primes with both EACH and a number are Between-Quantifier conditions. However, one of the central questions of this design is to ask whether there is more representational overlap between two quantifiers that may share a universal quantifier denotation (e.g. EVERY and EACH), as opposed to one universal quantifier and one existential with a specified set cardinality (e.g. EVERY and THREE). Looking only within the Between-Quantifier conditions, we separate universal primes EACH, EVERY, and ALL from non-universal number primes. Taking only the universal targets (EACH, EVERY, and ALL), we fit another logistic mixed effects model with fixed effects of Prime Scope and the factor of Universal-vs-Number-Prime, as well as their interaction. If, for example, there was a priming effect of EACH on EVERY, but not of THREE on EVERY, we would expect to see a significant interaction term. However, there were no significant interactions for any of the universal target quantifiers: EACH, EVERY, or ALL.

Across the four target types, there was consistent scopal priming when and only when the prime and target shared the same quantifier. This pattern seems to suggest that the operations that construct logical form are in some way specific to the individual quantifiers that are being manipulated. For example, we would expect a pattern like this if a different Quantifier Raising operation was used in each case or if Logical Form templates were stored in a lexicalized grammar.

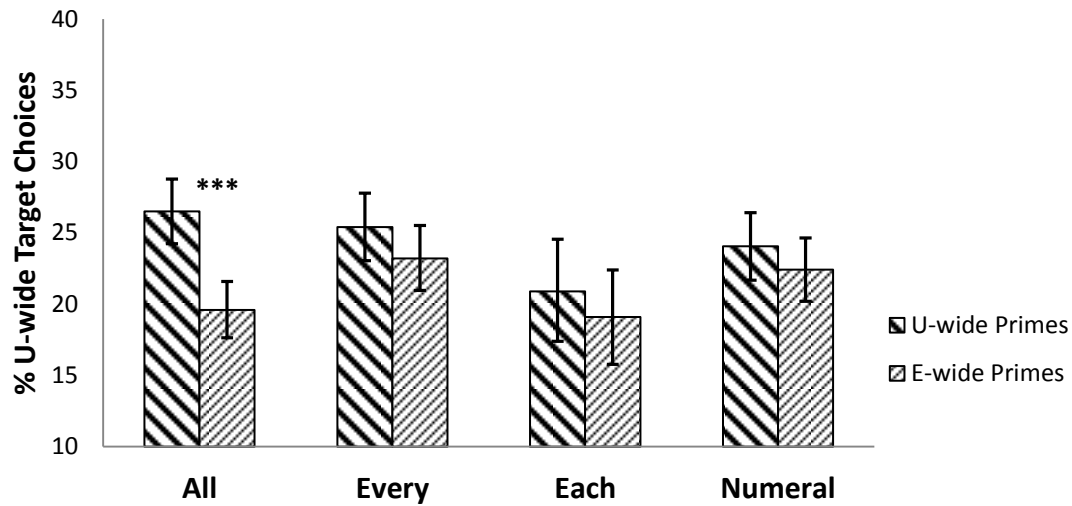
a) Every targets



b) Each targets



c) All targets



d) Number targets

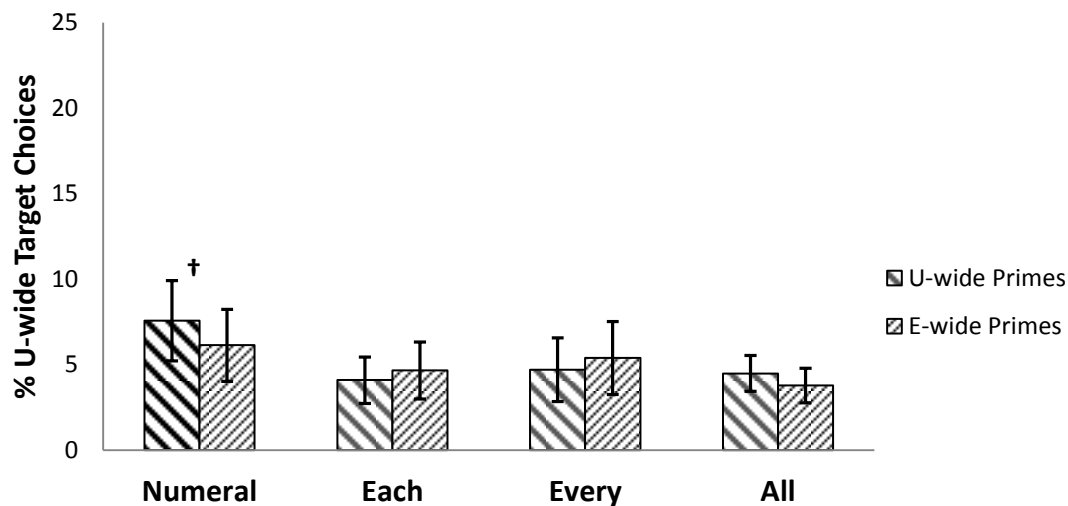


Fig. 4.4: Participants' rate of U-wide choices on ambiguous target trials in Experiments 2, separated by whether the target quantifier was (a) EVERY (b) EACH (c) ALL or (d) numbers. Note the different scales on the Y-axes between panels. Responses are broken down by whether the target trial was preceded by a U-wide or E-wide prime trial. The leftmost pair of bars in each graph shows the within-quantifier priming condition in that experiment, where the quantifier word is the same in both prime and target trials. The other pairs of bars show the between-quantifier conditions, with other quantifier words used in the prime sentences.

4.3.2.3. Context effects: Effects of Prime Quantifier independent of Prime Scope

In addition to the priming effects of a prime on the target that follows it, seeing sentences with the same prime quantifier repeated throughout the experiment may have a contextual effect on target choice, independent of the prime scope in a specific prime-target pair. As Experiment 1 shows, some quantifiers are strongly biased toward a U-wide interpretation (EACH being the most biased) and others are heavily E-wide biased (the numbers most so). Despite a participants' preferred reading, half of the prime trials a participant sees always force a U-wide reading, and the other half, an E-wide reading. This gradually gives participants increasing evidence that the normally dispreferred reading may be an acceptable one for that quantifier, and possibly for other quantifiers that are seen as similar. Thus, if participants repeatedly see a quantifier that they would normally assign either a U-wide or E-wide interpretation in the context of prime trials where the only available reading is the dispreferred one, they may adjust their expectations over time, resulting in a weaker bias on subsequent target trials containing that quantifier. If two different quantifiers are biased in the same direction, this effect may extend from one to the other in Between-Quantifier cases. In all cases, we would expect such a context effect to push target trial responses away from the baseline for that individual quantifier and closer to chance.

4.3.2.4. Within-Quantifier Context Effects

In the Within-Quantifier conditions, participants see the same quantifier word on subsequent trials – within each prime-target pair – as well as repeated through the experiment. Since each quantifier has a strong baseline bias, experience with primes that only match this bias half of the time could shift performance toward chance. For each target quantifier, we constructed a separate model comparing the baseline data from Experiment 1 to the Within-Quantifier priming condition for that word in Experiment 2. The models included a between-

subjects two-level variable of Prime Presence as the main predictor of interest (for example, the EVERY baseline from Experiment 1 compared to the EVERY-to-EVERY priming condition in Experiment 2).

There was a significant effect of Prime Presence for every quantifier, except in the ALL-to-ALL condition ($z=1.61$, $p=0.1$), suggesting that participants were sensitive to the presence of primes where they were forced to pick a dispreferred interpretation (see Fig. 4.5). We replicated the ALL-to-ALL Within-Quantifier condition with an additional 256 participants, of which 235 were included, after the usual exclusions. Including both the original run and replication data, we first fit a logistic mixed-effects model to look for the effect of Prime Type in this condition, but adding a two-level fixed effect of replication (Replication vs. Original Run) and found no effect of replication, nor any interaction between prime type and replication. In this data, we replicated a highly significant effect of prime type (U-wide or E-wide), with $p<0.0001$ (see Sec 3.2.1.). Pooling both the replication and original data, we fit a model comparing the Within-Quantifier ALL condition to the ALL baseline from Experiment 1. Including the additional replication data, we now found a significant effect of Prime Presence ($z=2.51$, $p=0.01$), indicating a significant difference between the Within-Quantifier priming ALL-to-ALL condition and the ALL baseline condition. Consistent with the results for the other quantifiers, the priming condition was closer to chance (23% mean U-wide responses) than the baseline condition (21% U-wide responses). Given this pattern of results, we conclude that the original failure to find an effect in the ALL-to-ALL condition was a false negative, and that there is a significant context effect between all Within-Quantifier conditions and their respective baselines.

This pattern of results confirms that all of the quantifiers we tested allow for both readings. While all quantifiers were biased toward one of the readings, the context effect (as well

as the priming effect) show that experience with multiple readings can result in systematic shifts in how these ambiguous sentences are interpreted.

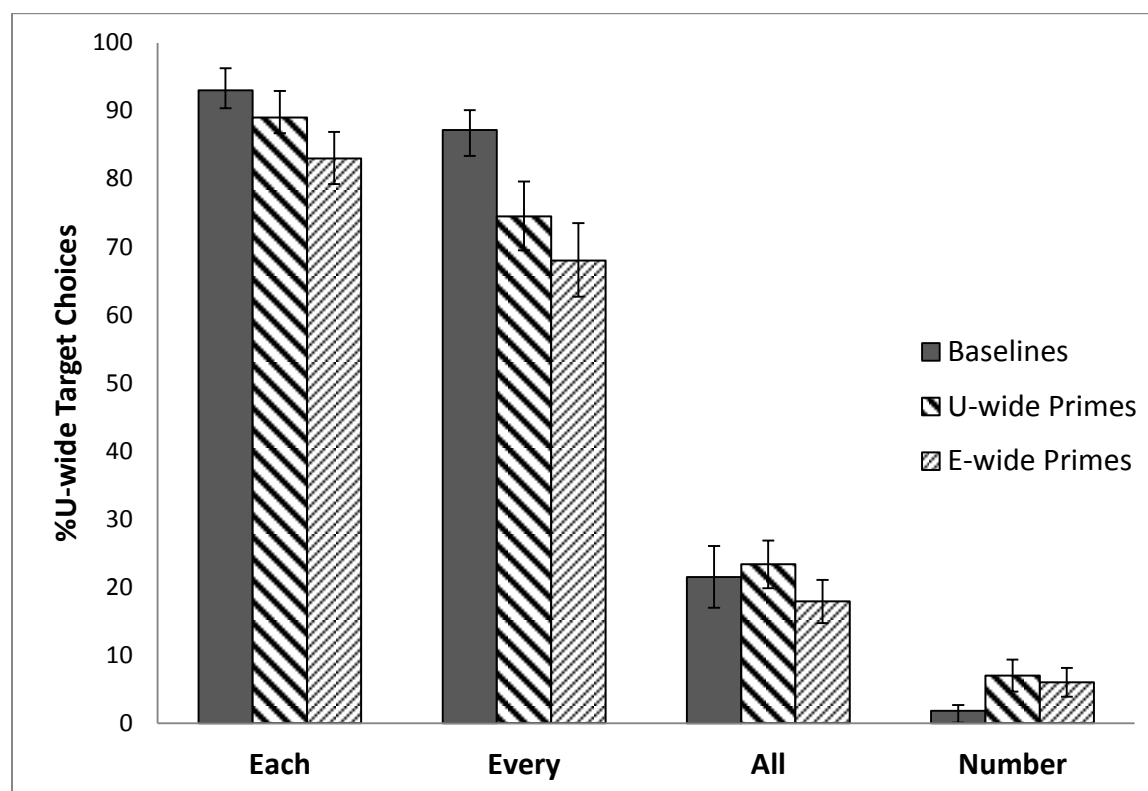


Fig. 4.5: The Within-Quantifier conditions of Experiment 2 alongside the baselines from Experiment 1. Error bars indicate 95% confidence intervals, averaged across items with subjects as the random variable.

4.3.2.5. Between-Quantifier Context Effects

Given that distributional information about a quantifier's scope preference is accumulated and stored, we can ask which sets of quantifiers this information is aggregated over. Does information that changes the bias of one quantifier change the bias of the other as well? It could be that probabilities get updated only for a specific quantifier word, in which case we would expect to see main effects of Prime Quantifier in the Within-Quantifier conditions, but not Between-Quantifier condition, just as with the priming effects. Alternatively, it could be that two or more quantifiers form a single class, such that evidence that one quantifier is more or less biased than expected would shift the bias of another quantifier in the same class. Finally it is

possible that these probabilities are aggregated across all of the quantifiers as a single class, in which case there should be effects of Prime Quantifier in all of the Between-Quantifier conditions.

The relevant statistical test of this hypothesis – a test for the main effect of Prime Quantifier for a given target, independent of the Prime Scope – were hidden away in the Between-Quantifier analyses we discussed in Sec 3.2.2 above. Recall that in this analysis we constructed four separate models, one for each target quantifier, each of which included the following predictors: a three-level fixed effect of Prime Quantifier (capturing the three different Between Quantifier conditions), a two-level fixed effect of Prime Scope, and their interaction. Across these four models, we find only two significant main effects of Prime Quantifier (out of a total of 12 possible effects). In the model where EVERY is the target, we find a main effect of the condition where EACH is the prime ($z=-3.49$, $p=0.0005$), and in the model where EACH is the target, we find a main effect of the condition where EVERY is the prime ($z=-2.96$, $p=0.003$). To follow up on these effects, we compared these conditions to respective baselines for those target quantifiers (from Experiment 1). In a model including just the EVERY baseline condition from Experiment 1 and the EACH-to-EVERY priming condition from Experiment 2, we find a highly significant effect of Prime Presence ($z=-7.41$; $p<0.0001$). Similarly, in a model including the EACH baseline condition from Experiment 1 and the EVERY-to-EACH priming condition from Experiment 2, we find a significant effect of Prime Presence ($z=-3.19$; $p=0.001$). These effects are parallel to the Within-Quantifier context effects: the presence of a prime pushes target responses down closer to 50-50 and away from the quantifier's bias.

We interpret these effects as evidence that participants adjust their expectations about the scope bias of EACH based on distributional information about EVERY, and vice versa. Why does

this information get aggregated across this pair of quantifiers and not across other pairs? Perhaps there is some semantic property that is shared by *EVERY* and *EACH*, making them more similar to one another than they are to *ALL* or to numbers (see Champollion, 2010; Steedman, 2012, and Sec. 4.6.3. and 4.6.4. for theoretical considerations along these lines). Note, however, that any similarity of this kind is apparently not sufficient to allow scopal priming between *EVERY* and *EACH*. The systematic lack of Between-Quantifier priming effects between *EACH* and *EVERY* and the consistent presence of Between-Quantifier context effects for this pair suggests that there are at least two distinct types of representations or processes involved in LF construction (operating over different time scales), which our theoretical accounts should address. Constructing an LF online is a process that draws on prior beliefs about the scope bias of the quantifiers involved. The process that updates these prior beliefs appears to be sensitive to a representational similarity between *EACH* and *EVERY*. In contrast, the process that constructs the representation of quantifier scope (presumably based on the instructions provided by these priors) seems to treat these quantifiers as wholly distinct.

4.3.2.6. Addressing deflationary accounts of the priming effects

The pattern of findings reported above rules out several deflationary accounts of the locus of priming. First, given that we use a picture-choice task for both primes and targets, one deflationary possibility is that any priming effect might be due to greater visual similarity between a U-wide target picture and a U-wide prime – rather than an E-wide prime – picture. All U-wide pictures have pairs of agents acting on themes, while all E-wide pictures show many agents acting on a single theme. Similarly, the U-wide and E-wide pictures differ in the kinds of events they depict. In the U-wide picture, there are multiple events – each agent acts separately on a different theme. The E-wide picture depicts a single event, where all of the agents act

together on the same theme. Another deflationary possibility is that the priming effect has little to do with linguistic representations, and occurs instead at the level of event representations.

However, the significant interaction between the Within- and Between-Quantifier conditions and the lack of priming Between-Quantifiers provide strong evidence that picture similarity and event similarity are not responsible for this form of priming. Not only are the pictures and event structures the same in the Between-Quantifier conditions that do not prime as in the Within-Quantifier conditions that do, but the prime sentences in the Between-Quantifier conditions are ones that are capable of priming – the very same prime sentences that produce a priming effect in their respective Within-Quantifier conditions.²¹

Our findings demonstrate that scopal priming is mediated by linguistic representations: it depends on the identity of the quantifier and not just the picture or the event structure. Any deflationary account of priming would have to capitalize on the distinction between the conditions where priming is observed (Within Quantifier) and the conditions where it is absent (Between Quantifier). For example, the adjacent prime and target sentences in the Within-Quantifier cases are more similar to one another, at the phonological and lexical levels, than the corresponding sentences in the Between-Quantifier cases, simply because they share an additional word. Maybe this global increase in sentence similarity leads participants to draw comparisons between the two sentences or to infer that similar responses are expected. Experiment 3 explores whether reducing the similarity of the prime and target sentences, without changing the quantifier, will decrease or eliminate scopal priming. In Experiment 2 (and in R&P)

²¹ Raffray and Pickering (2010, Experiment 4) argue against an event- or picture-based interpretation of the priming effect, using a control condition that pairs the same pictures with generic sentences not containing quantifiers (e.g. “Kids climb trees”). They find no priming effect, but an overall inflated rate of target U-wide responding relative to sentences quantified with EVERY. This suggests participants may interpret such generics as closely matching U-wide pictures, and so the lack of priming may be specific to the U-wide reading of generics overriding any picture priming effect, rather than a demonstration that there is no such effect when the sentences are scopally ambiguous.

the primes and targets always had the same verb. In Experiment 3, the primes and targets have different verbs (but the same quantifier). We chose to manipulate similarity by changing the verb because verbs define the event type and are thus likely to affect perceived similarity.

Furthermore, while verbs have no privileged role in interpreting quantifiers, they do play a critical role in language processing. For example, in studies of syntactic priming (the priming of argument structure alternations) effects are often larger when two sentences share a verb than when they do not (see e.g., Pickering & Branigan, 1998). If scopal priming depends on mere similarity, it should be reduced or absent in this Between-Verb study. In contrast, if the effects in Experiment 2 are due to quantifier specificity in the operations that build Logical Form, then priming in this Between-Verb study should be equal to the corresponding condition of Experiment 2.

4.4. Experiment 3: Between-Verb priming

4.4.1. Method

4.4.1.1. Participants

An additional 128 participants were recruited via Amazon Mechanical Turk. The exclusion criteria were the same as for the previous experiments. After these exclusions were applied, 107 participants remained (M Age: 31.7, Range: 18-64).

4.4.1.2. Stimuli and Procedure

Experiment 3 used the same stimuli and procedure as the Within-Quantifier condition with EVERY in both prime and target sentences, with the only difference being in the pairing of prime-target trials; prime-target pairs were scrambled so that target trials no longer followed

prime trials containing the same verb (see Fig. 4.6 for an example).

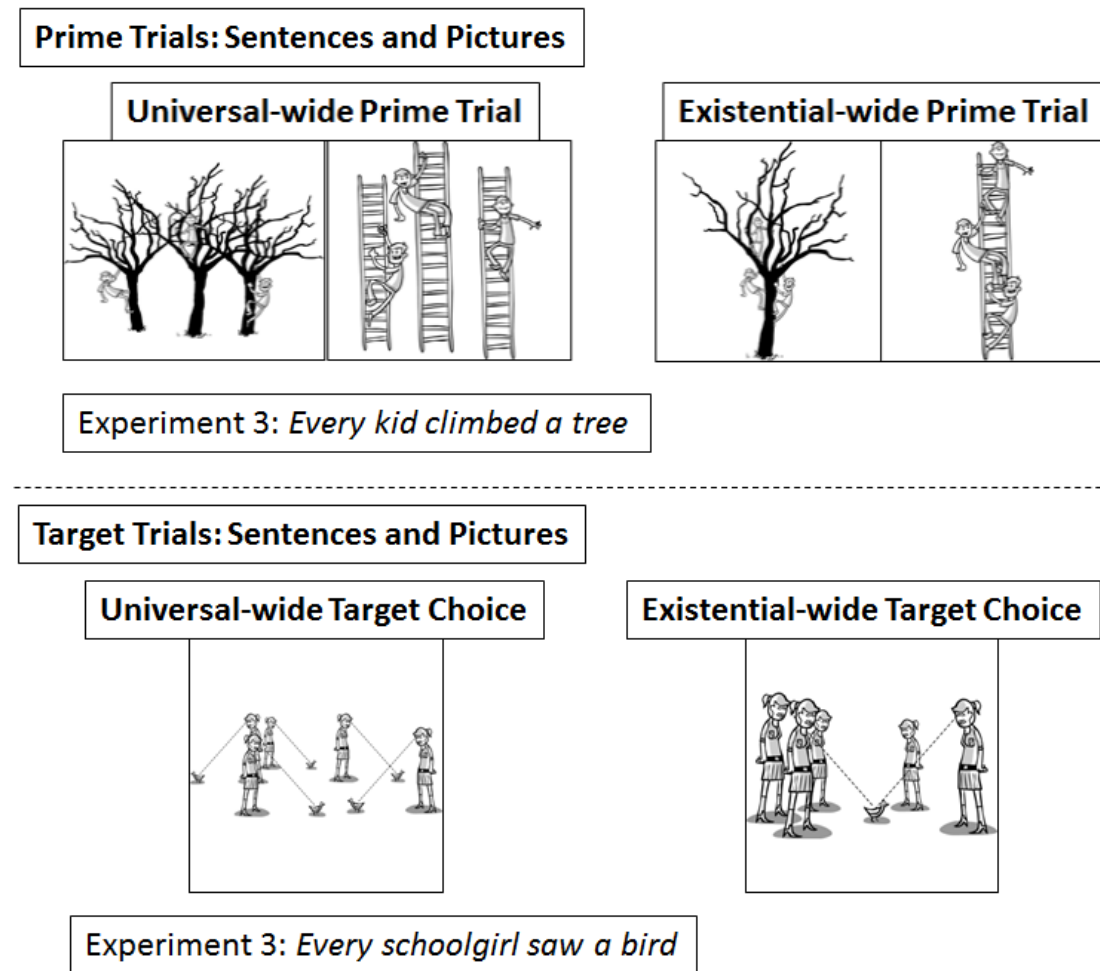


Fig. 4.6: Sample prime and target items for Experiment 3. Prime trials (top) always directly preceded target trials (bottom). There were two types of prime trials – U-wide and E-wide. The correct picture choice on the prime trial depicted the corresponding reading of the sentence while the incorrect choice pictured a noun not present in the sentence. The pairings of prime and target pairs was reshuffled in Experiment 3 relative to Experiment 2, so that prime and target sentences no longer shared their verb.

4.4.2. Results and Discussion

The data from target trials was analyzed using the same type of logistic mixed-effects model as Experiment 2, with a maximal random effects structure. We modeled response type on the target trials, with random slopes (relative to Prime Scope) and intercepts for subjects and

items and Prime Scope as a fixed factor. We found a main effect of Prime Scope in this Between-Verb condition ($z=3.11$; $p=0.002$).

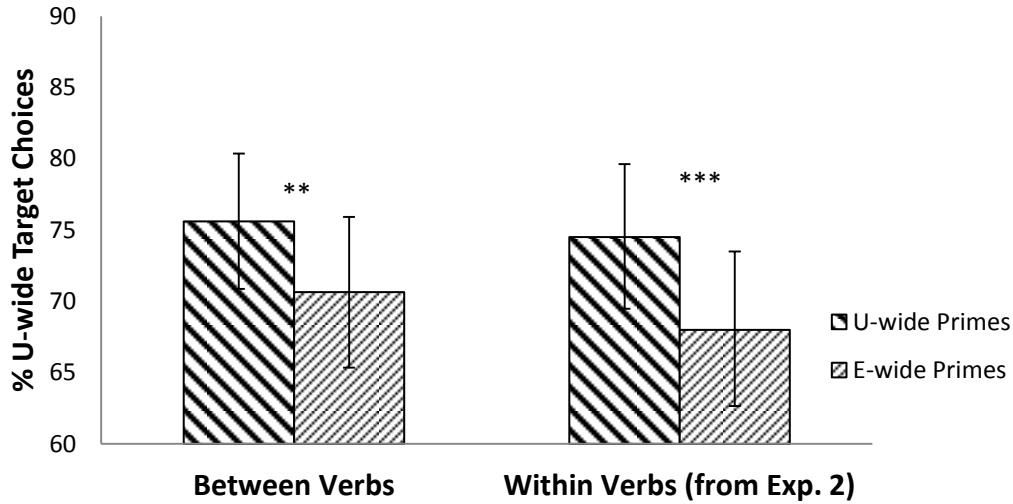


Fig. 4.7: Participants' rate of U-wide choices on ambiguous target trials in Experiment 3. Target trial responses are broken down by whether the target trial was preceded by a U-wide or E-wide prime trial. The left pair of bars shows the data from the Every-to-Every condition in Experiment 2. Error bars indicate 95% confidence intervals, averaged across items with subjects as the random variable.

To compare the Between-Verb results from Experiment 3 to the equivalent Within-Verb condition of Experiment 2, another model included the EVERY-to-EVERY Within-Quantifier condition, including the same random effects, as well as the fixed effects of Prime Scope, Verb Overlap (Between- vs. Within-Verb) and their interaction. This model found no significant interaction ($z=-0.8$; $p=0.42$) and no main effect of Verb Overlap ($z=0.35$; $p=0.73$), but a highly significant main effect of Prime Type ($z=4.31$, $p<0.0001$). Figure 4.7 shows the rate of U-wide responding in Experiment 3 and the EVERY-to-EVERY condition of Experiment 2, which are similar. These results demonstrate that priming is not dependent on global similarity. Changing the verb between prime and target had no effect on the magnitude of scopal priming. In contrast, changing the quantifier in Experiment 2 eliminated scopal priming.

There are, however, two different ways in which priming could be quantifier-specific. First, priming could be tied to the phonological label of the quantifier. For example, each phonological label could have two semantic entries, one of which contains a type-shifting, QR, or some other operation to assign the quantifier wide scope, while the other does not. Thus priming of scope would occur only through a given phonological label. The second possibility is that quantifier specificity could be tied to substantive differences in the meanings of the operators. While scope ambiguity is a general property of quantification, some theorists have invoked distinct combinatorial structures and mechanisms to explain both general combinatorial and specific scopal properties of EVERY, EACH and ALL (Steedman, 2012; Beghelli & Stowell, 1997; Champollion, 2010; see the General Discussion). Perhaps it is the differences between these meanings that are responsible for the specificity of the scopal priming effects.

To disentangle the effects of phonological form from shared meaning, in Experiment 4 we look for priming when a different number is used in the prime sentence than in the target sentence (THREE, FOUR or FIVE). Although there are many different analyses of the combinatorial semantics of number words (see, for example, Steedman, 2012; Heim and Kratzer, 1998; Beghelli and Stowell, 1997), within a given theory, different numbers always have identical combinatorial properties. The only differences between them are in the cardinality of the set they pick out. Thus, if different number words do not prime each other, it would suggest that the priming effect is mediated by their phonological labels. If sentences containing one number word do prime sentences containing a different one, then it would show that priming can occur independent of the phonological label. That would in turn suggest that the lack of priming between quantifiers in Experiment 2 is a sign of differences in their meanings – specifically, in the semantic mechanisms of scope construction of each quantifier.

4.5. Experiment 4: Between numbers

4.5.1. Method

4.5.1.1. Participants

An additional 128 participants were recruited. After applying the same exclusion criteria as in the preceding experiments, 112 participants (M Age: 28.5, Range: 18-64) were included in the final analysis.

4.5.1.2. Stimuli and Procedure

The stimuli and procedure for Experiment 4 were identical to the Number-to-Number condition in Experiment 2, except that additional subject entities in the prime pictures were either added or removed from the prime pictures, so that the number of subject entities in the prime trial mismatched the number in the target trial (see Fig. 4.3). The target trials were unchanged for the sake of comparability with Experiment 2. For example, if the target trial presented a choice between two pictures – one with three hikers climbing one hill (E-wide), and one with three hiker-hill pairs (U-wide), the E-wide prime trial in Experiment 4 presented a choice between four kids climbing one tree or four kids climbing one ladder, and the U-wide prime trial presented a choice between four kid-tree pairs and four kid-ladder pairs. The prime sentence would correspondingly be, “Four kids climbed a tree”, while the target sentence would remain as before, “Three hikers climbed a hill”.

4.5.2. Results and Discussion

The data from target trials was analyzed using the same type of logistic mixed-effects model as the previous two experiments, with a maximal random effects structure. We again modeled Response Type on the target trials, with random slopes and intercepts for subjects and items, and with Prime Scope as a fixed factor. There was a significant effect of Prime Scope

($z=2.07$; $p=0.04$). In a second model including the Between-Number data from Experiment 4 and the Number-to-Number Within-Quantifier condition from Experiment 2 (where the number words used in a given prime-target pair were always the same), we added the fixed effect of Number Overlap and its interaction with Prime Scope. In this model, there was no main effect of Number Overlap ($z=-0.85$; $p=0.39$) and no significant interaction with Prime Scope ($z=0.12$; $p=0.9$). There was a significant main effect of Prime Scope ($z=2.26$; $p=0.02$). Experiment 4 demonstrates that priming does not depend on whether two quantifiers share the same phonological label. Priming from THREE to FIVE was as robust as priming from THREE to THREE. We discuss the significance of these findings in relation to linguistic theory in the General Discussion below (see 4.6.2).

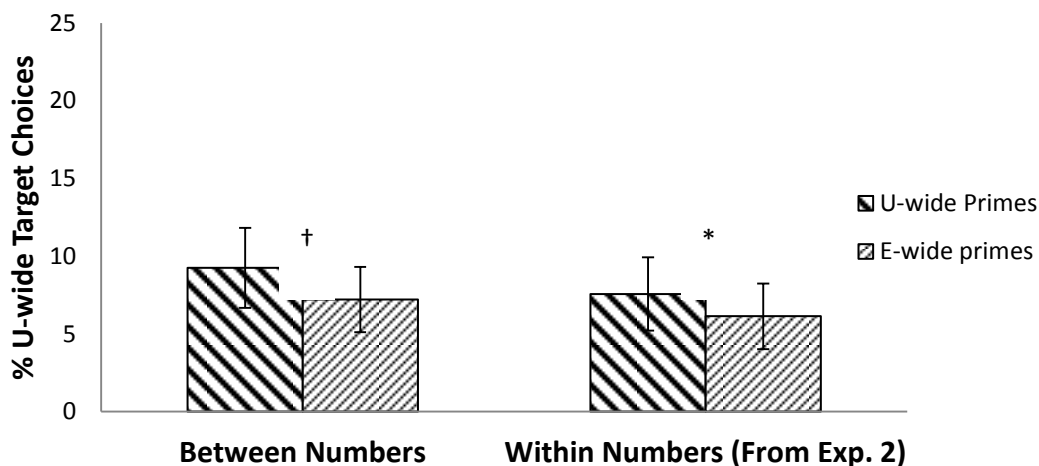


Fig. 4.8: The left pair of bars (Between-Numbers) shows participants' rate of U-wide choices on ambiguous target trials in Experiment 4, while the right pair shows participants' U-wide response rate on the Number-to-Number Within-Quantifier condition of Experiment 2 (Within-Numbers). Target trial responses are broken down by whether the target trial was preceded by a U-wide or E-wide prime trial. Error bars indicate 95% confidence intervals, averaged across items with subjects as the random variable.

4.6. General Discussion

The four experiments described above shed light both on the mechanisms of scope ambiguity resolution, and on the representation of quantifiers. In Experiment 1 we find large baseline differences in scope-taking behavior between the quantifiers EVERY, EACH, ALL, and numbers. Experiment 2 uses a priming paradigm to examine how forcing the resolution of scope ambiguity in one sentence influences the interpretation of a subsequent one. We find that forcing a U-wide or E-wide resolution of the prime sentence influences how participants resolve a subsequent sentence, but only when the subject quantifier word in the prime and target sentences is the same. Changing the quantifier word eliminates priming. Separate from priming, we find a context effect that occurs when the quantifiers in the primes and the targets are the same, and between the quantifiers EACH and EVERY. In these cases, repeated experience with primes that illustrate both scope readings shifts our participants' interpretation of the target sentences in the direction of chance (relative to the baseline) suggesting that their expectations for the scope of target quantifier has changed. Experiment 3 shows that the difference in priming between the Within- and Between-Quantifier conditions of Experiment 2 cannot be due to overall similarity, since the priming effect is still present when the prime and target sentences have different verbs. Experiment 4 demonstrates that Between-Quantifier priming is possible, when the quantifiers are different numbers, indicating that priming does not depend solely on the phonological label associated with the quantifier, but can occur when two different quantifiers have meanings that are sufficiently similar. These findings have implications: for our understanding of the mechanisms behind quantifier-scope phenomena, for the question of whether and how the three universal quantifiers differ in their meanings, and for the interface between language and conceptual representations.

4.6.1. Lexical versus structural factors in preferred scope reading

Although many theorists have reported intuitive judgments about differences in scope preferences between quantifiers (e.g. Quine, 1960; Vendler, 1967; Ioup, 1975), the present studies provide quantitative evidence for variation between the scope-taking preferences of different quantifiers, lending empirical support to a specific ordering of scope preference across the quantifiers (EACH>EVERY>ALL>numbers). The differences that we observed across the quantifiers in Experiment 1 are dramatic: participants chose the U-wide reading 93% of the time for EACH, but only 2% of the time for the numbers. Such stark differences are easily explained if quantifier scope depends largely on lexical factors, but they are difficult to reconcile with accounts on which the resolution of scope ambiguity is driven by structural factors such as linear order (Johnson-Laird, 1969; Lakoff, 1971; Kroch, 1974; Fodor, 1982; Bunt, 1985), surface syntactic c-command (Jackendoff, 1972; VanLehn, 1978; May, 1985; Lidz and Musolino, 2002; Anderson, 2004), or a thematic hierarchy in which quantified agents take scope over quantified themes (Ioup, 1975; Jackendoff, 1972; Grimshaw, 1990; Kurtzmann & MacDonald, 1993). Our findings show that, to the extent that these other factors play a role (and they probably do; see Kurtzmann & MacDonald, 1993) they can be overridden by the lexical or conceptual properties of individual quantifiers.

Why then have prior studies failed to observe these strong lexical biases? We suspect that it is because they have not looked for them. Most of the prior studies have focused on one pair of quantifiers (EVERY and A) and thus were not designed to compare the biases of different quantifiers. Because most of the studies looking at structural factors did not employ the most strongly biased quantifiers (the numbers and EACH), they may have overestimated the effects of structural variables. Where lexical constraints are weak the effects of syntactic or thematic

constraints may be clearest, but where lexical constraints are strong they may overwhelm these other biases. Moreover, recent work by Chemla and Bott (2015) shows less support for structural factors even when looking only at the scopal relationship between EVERY and A. Using a picture priming paradigm, they find that assigning EVERY a wide scope reading in one sentence primes EVERY to take wide scope in a subsequent sentence, even when linear, thematic and syntactic structural factors are specifically stacked in the other direction.

We find not only that different quantifiers have different preferred scope biases, but that these biases can shift when seeing that a quantifier may appear in its typically dispreferred scope position. The context effects we find in Experiment 2 (separate from the trial-by-trial priming effects) suggest that knowledge of a quantifier's preferred scope can be updated over the time it takes to complete the experiment. While this information is aggregated across trials within a single quantifier, we also find that information about the possible readings for EACH affects the scope bias of EVERY and vice versa, suggesting that updating of scope bias may involve some shared semantic properties between just these two quantifiers. In contrast, the online assignment of relative scope (which draws on quantifier biases as one input, and which is reflected in the trial-by-trial priming effects) is not sensitive to these same features.

That quantifier words vary so widely in their preferred scope readings suggests people store some information on how to construct combinatorial meaning in the lexical entries of different quantifiers. But what are these entries like? Our findings point to a distinction between two different levels of representation for quantifier meaning – the conceptual content of a quantifier word as one level and its combinatorial properties as another.

4.6.2. Combinatorial semantics versus conceptual content

An assumption commonly made in the linguistics literature (e.g. Heim & Kratzer, 1998) is that the combinatorial semantic system which produces complex sentence meanings from individual words is distinct from the conceptual system which stores each individual lexical item's content. This is why semantic denotations for many content words are conspicuously vacuous. For example, the denotation for SWIM typically looks something like this:

$$[\text{swims}] := \lambda x.x \text{ SWIMS}$$

This denotation specifies the way in which the word “swims” combines with other words. SWIMS is a function that maps individuals to sets of individuals. Specifically, it maps entities (such as John, who might feature in the sentence, “John swims”) to a set (in this case, the set of swimming things). But what constitutes a swimming thing? The combinatorial semantics do not say. The concept SWIMS is invoked on the right hand side of the equation only as a pointer to some content, which is assumed to be stored separately from the combinatorial properties of the word (perhaps in a single amodal conceptual store or perhaps as a distributed representation that is more tightly linked to perception and action). In just the same way, the differences in meaning between different number quantifiers are assumed to be stored in a separate conceptual system, with the semantic combinatorial system at LF providing pointers to the relevant conceptual entries. The combinatorial semantics of any number word specify that There exists a set of M with cardinality N, where the value of M is filled in by the noun being quantified, and the value of N is given by the number word. The conceptual content of any specific cardinality N – its intensional and extensional properties – would be stored in a separate conceptual inventory.

The pattern of priming effects that we find is quantifier-specific in a way that closely reflects this division of labor. That we find priming effects between different numbers (THREE,

FOUR and FIVE), but not between different universal quantifiers (EACH, EVERY and ALL) supports the view that the former differ only in which concept they point to, while the latter differ in terms of their combinatorial semantic structure. The differences between the meanings of THREE, FOUR, and FIVE are handled by a different set of cognitive systems than the ones responsible for the construction of LF and the resolution of scope ambiguity.²² EACH, EVERY and ALL, on the other hand, do appear to differ in terms of their combinatorial properties.

While the division of labor between conceptual content and combinatorial properties is common to all linguistic theories of meaning, there is a wide array of specific theories of LF construction and scope assignment. We find that the mechanism that constructs LF is quantifier-specific rather than abstract, that different quantifiers prefer very different scope assignments, and that these preferences can shift depending on distributional information about possible readings for each quantifier, and between EACH and EVERY. While these findings provide new constraints on all theories of LF, they also fit more naturally with some accounts than with others.

4.6.3. A range of quantifier scope theories

For our purposes, theories of quantifier scope can be divided along a continuum, depending on how much they favor “lumping” or “splitting” of scoping mechanisms across different quantifiers. Theories on the lumping end posit a common mechanism responsible for all scope phenomena, while those on the splitting end hold that all quantifiers differ from each other in their scopal mechanisms. There are also many intermediate positions along the continuum, positing shared properties for some, but not all quantifiers. The extreme lumping end of the continuum, in particular, is densely populated. Theories at this end range from positing that all

²² See for example, Dehaene (1997), LeCorre and Carey (2007), for discussion of how the meanings of different number words are mapped to different representations in the Approximate Number System and other non-linguistic representational structures.

quantifiers share common syntactic operations, like Quantifier Raising, or QR (May, 1977; 1985) and quantifying-in (Montague, 1972; Rodman, 1976), to semantic mechanisms, like various forms of type-shifting; e.g. Barker, 2002; Hendriks, 1988) or Cooper Storage (Cooper, 1983; Keller, 1988).²³ All quantifier words in these theories are treated as Generalized Quantifiers in the sense first introduced by Montague (1972), with semantic denotations bearing the logical force of quantifiers in first-order predicate logic. Syntactically, too, all are members of the same category.

The combination of within-quantifier priming effects for all quantifiers and the lack of any priming effects between quantifiers cannot be explained by any single abstract operation acting on all of the different quantifiers during the online construction of meaning. It suggests instead that the operation or semantic feature responsible for scoping behavior is individualized by quantifier, possibly stored as part of each quantifier's lexical entry. Our findings are unexpected from the perspective of extreme lumping theories – the dominant framework in formal semantics for many decades – because, if the locus of priming was a single mechanism, it should have led to priming between all pairs of quantifiers.

Theories that favor some degree of splitting take as their central fact that quantifiers differ in their patterns of permitted scope-taking behavior. To give one common example, compare the scoping behavior of EVERY and A in relative clauses (Fodor & Sag, 1982). EVERY does not appear to be able to take wide scope when it is contained in a relative clause. Take the following sentence:

(5) A teacher heard the rumor that every student of mine was called before the dean

²³ See Von Stechow (1991) for discussion of how semantic and syntactic approaches in this class become more similar as they posit more powerful operations on either end.

There does not seem to be a U-wide reading where, for every student, there is a different teacher. In contrast, when A is within a relative clause instead of EVERY, it takes wide scope readily.

(6) Every teacher heard the rumor that a student of mine was called before the dean
Both readings of (6) are available – both a U-wide reading where every teacher heard about a different student, and the E-wide reading where every teacher heard about the same one.

To account for the finding that A can scope out of relative clauses while EVERY cannot, Fodor and Sag (1982) propose splitting the mechanisms by which these quantifiers take wide scope. They argue that the indefinite A has an alternative meaning that is referential and not quantificational, and that in its referential form, A can take scope freely anywhere in a sentence, even if it is contained in a relative clause.

While Fodor and Sag (1982) attempt to preserve a large degree of lumping by arguing that A is unique in having a non-quantificational reading and that all other quantifiers have a common mechanism for taking scope, later theorists have committed more strongly to the extreme splitting end of the continuum, positing different scoping mechanisms for different quantifiers (eg. VanLehn, 1978; Reinhart, 1997; Beghelli and Stowell, 1997; Winter, 2001; Champollion, 2010; Steedman, 2012 and many others). Perhaps the most detailed extreme splitting account is given by Beghelli and Stowell (1997), who argue that there is no single mechanism for LF construction and posit a fully lexicalized system of scope, with feature-based denotations of quantifiers in the lexicon and feature-dependent landing sites at LF. Quantifier-scope is determined by c-command relations in the resulting constructed LF, and scope ambiguities arise because some quantifiers have multiple features, with different ones being active in different sentences and contexts. One feature of A, if active, would land it at a site

higher than the site that attracts EACH, and the other would land it lower, corresponding to A taking either wide or narrow scope relative to EACH.

Although nothing in our study bears on the specifics of Beghelli and Stowell's (1997) proposal, our primary finding favors theories on the splitting end of the continuum. We find robust within-quantifier priming but no between-quantifier priming. We interpret this pattern as evidence that the procedures involved in constructing Logical Form during language comprehension are specific to individual quantifiers (or to narrow semantic classes of quantifiers, such as numbers). For example, when we construct a representation where EVERY takes wide scope over A, we may activate a representational chunk that captures this precise scopal relation. If we later encounter a sentence which contains both EVERY and A, we will be more likely to activate the same scopal structure than the alternative (see Sec. 4.3.2.1.). However, if we encounter a sentence with EACH and A, then neither of these representations will be relevant and no priming will occur (see Sec. 4.3.2.2.).

However, these studies also demonstrate that there are representational similarities across quantifiers, similarities which are not active on a trial-by-trial basis but can influence scope resolution on a longer time scale. Specifically, in Experiment 2, we observed context effects which were present both when the prime and target had the same quantifier and between primes or targets with EACH and those with EVERY. We interpret these context effects as the result of participants updating their prior beliefs about the target on the basis of the prime sentences. The fact that examples of EACH influence beliefs about EVERY (and vice versa) suggests that learners represent some commonality between them, a commonality that is not shared by ALL or the numerals. Steedman's (2012) book offers one possible explanation. He proposes that only the denotations of the quantifiers EVERY and EACH contain the universal logical operator, \forall . Other

quantifiers, like A, THREE and ALL, use entirely different formal combinatorial machinery (Skolem functions and constants; see also Reinhart, 1997; Winter, 2001). Along the continuum from lumping to splitting, Steedman's proposal lies in the middle. He groups EVERY and EACH into a single category, but splits these off from all the other quantifiers. As we mentioned, the absence of between-quantifier priming effects on a trial-by-trial basis gives no reason for grouping EACH with EVERY, or any of the other quantifiers together. If we assume, however, that the representational similarities posited by Steedman guide the rapid learning observed in the present study rather than online LF construction on a trial-by-trial basis, then Steedman's theory would predict context effects between the numbers and ALL, and between EVERY and EACH. We find one of these effects (the context effects between sentences with EVERY and EACH) but not the other (no effect between numbers and ALL). At first blush this suggests that Steedman's theory is right only in part. However, the predictions of such a theory depend on whether the different Skolem functions employed by various non-universal quantifiers are similar enough to each other to form a common basis for generalization. If they are not then the context effects may be fully consistent with this proposal.

4.6.4. Preferred versus possible scope readings

All of the theories described above, regardless of their position on the lumping-splitting continuum, account for the possible scope behavior of different quantifiers, but they are largely mum on explaining differences in preferred readings. Even theories towards the splitting end, which posit semantic differences between quantifiers (e.g. Beghelli and Stowell, 1997), seek to explain which readings are allowed and not why some of the allowable readings are preferred to others. We find that all of the quantifiers we tested can take either wide or narrow scope relative to A and that even extremely dispreferred readings can become more available due to context and

priming effects. Most linguistic theories of quantifier scope assume that differences in preferred readings are due to the effects of discourse context, and so foist off the explanation of these differences to the domains of pragmatics and domain-general, extra-linguistic cognition.

The results of Experiment 1 challenge this assumption. All of our target sentence were presented in exactly the same context regardless of the quantifier used: the filler trials, the instructions, the specific pictures presented, and the other words in the target sentences were identical across conditions. Only the quantifier in the subject noun phrase varied. Nevertheless, we found systematic baseline differences in the scopal preferences of the quantifiers we tested: while EVERY and EACH strongly preferred to take wide scope, ALL and the number quantifiers preferred to take narrow scope. In fact, every quantifier in our study was reliably different from each of the others. Since these differences cannot be chalked up to different pragmatic or discourse-level effects, they must be attributable to some lexically encoded difference between the quantifiers, based either on their semantic content or on their patterns of use.

One factor that is likely to play a role is the semantic feature of Quantifier Distributivity, which is posited to be an inherent part of the meaning of some quantifiers but optional or absent in others (see Roberts, 1987; Beghelli and Stowell, 1997; Tunstall, 1998; Winter, 2001; Champolion, 2010, among others). A distributive quantifier is one that forces each atom in the quantified noun set to have the predicated property applied to it separately. For example, EACH is argued to be necessarily distributive and thus unacceptable in contexts where the predicate is necessarily collective, Dowty (1987) argues that this accounts for the contrast between (7a) and (7b).

(7a) *Each boy gathered in the yard

(7b) All of the boys gathered in the yard

Dowty argues that (7a) is ungrammatical because GATHER requires a collective noun, and EACH can only provide individual atoms rather than collections. ALL, on the other hand, is able to take the entire set of boys as a collection (but see Champollion, 2010). This distinction between distributive and non-distributive quantifiers seems to pattern with the preferences we saw in scope assignment: EACH is typically argued to be mandatorily distributive and showed the greatest preference for a U-wide scope reading; EVERY is often argued to be at least optionally distributive (Tunstall, 1998) and had a robust but less categorical U-wide preference; ALL, which seems to resist the distributive reading, showed a systematic E-wide preference. If EACH and EVERY both carry the distributivity feature, that could also account for the context effect we find in Experiment 2, where accumulated information about possible readings for EACH gradually affects the preferred scope bias of EVERY, and vice versa. But it is critical to note that differences in distributivity alone cannot directly predict performance in this study. While the E-wide pictures in our study might seem more "collective" than the U-wide pictures, both types of stimuli depict semantically distributive events. In both the U- and E-wide pictures, each subject is acting independently on its theme (e.g. each individual kid is climbing a tree in both the E-wide picture and in the U-wide picture). Thus we would need a hypothesis that connected distributivity both to online scope assignment and to the updating of quantifier scope bias to account for this data pattern. At present, we are not aware of any theoretical account of this kind.²⁴ The investigation of a principled connection between this feature of quantifier meaning and the quantifiers' scope behavior is a promising target for future research in both semantic theory and psycholinguistics.

²⁴ Beghelli and Stowell (1997) do link distributivity to the possible scope assignments of EACH and EVERY –these quantifiers can have a [+Dist] feature determines their landing site at LF. But they do not provide an account of why this would cause EACH and EVERY to preferentially take scope over A, since A can attach to a landing site either above or below DistP at LF. In other words, their account, like most others in the literature on scope, seeks to account for possible rather than preferred scope readings, and does not explicitly connect the two issues.

4.7. Conclusion

In relating our psycholinguistic investigation to semantic theories of quantifiers and quantifier scope, we attempt to connect the two fields to the benefit of both. In the field of semantics, theorists strive to provide adequate model-theoretic descriptions of the combinatorial machinery of language. In psycholinguistics and psychology, experimentalists strive to identify the mental representations that implement a combinatorial system in both language and thought. The experiments described above form a case study where these two projects constrain each other, providing important data for future theories to contend with.

We find empirical justification for a central assumption in linguistic theory – that representations of conceptual content are distinct from representations of the combinatorial properties of concepts. While for the universal quantifiers EACH, EVERY and ALL, the scope assigned to a quantifier in one sentence only influences the next sentence if it has the same quantifier, sentences with different numbers (THREE, FOUR and FIVE) influence each other. This suggests that the combinatorial machinery that assigns scope and constructs LF is sensitive to some differences between quantifiers (as between the different universals), but not to the differences between number concepts.

We also find systematic differences across quantifiers with respect to the preferred scopal reading. These biases can be pushed around through experience with the dispreferred readings, providing further evidence that this is merely a preference rather than a hard difference in acceptability. These preferences, and their ability to change with experience, do not map cleanly to anything posited in current theories of quantifier scope. Future linguistic theories should seek to explain the representational basis of this learning (what we are counting over when we update our priors) and how it relates to the mechanisms underlying quantifier scope assignment and LF

construction. A central unresolved question is why this updating process treats EACH and EVERY (but not ALL) as belonging to the same class. While we cannot neatly link these data patterns to existing semantic theories, we see these findings as part of a dialog between psycholinguistics and formal semantics which is still in its earliest stages. We are seeking out the phenomena that will drive future work, challenge our current theories and build a broader bridge between these fields.

Chapter 5:

Conclusion

5.1. The development of negation

The three papers in this dissertation present separate case studies in the development and structure of combinatorial thought. Each asks different questions, looking at different populations (infants, toddlers, adults). In the first paper we find evidence that infants represent other agents as having goals to approach an object, but not to avoid one, when given comparable amounts and types of information about each goal. Avoidance goals could be represented in one of two ways: either using a separate representation that is not compositionally related to approach, or composing approach with a mental symbol that has the content of logical negation, to form not-approach. Either one of these representations would be independently sufficient to represent an avoidance goal, and infants' failure suggests that neither type of representation is available to them. Since there is evidence, both from our task and from the literature on infants' goal representations (Woodward, 1998; 1999), of their representing approach, infants' failure to represent not-approach points either to their not having a negation operator or their being unable to combine that operator with approach, perhaps because the approach representation is not a concept – that is, does not have the right format for productive, meaningful combination with other concepts.

In the second paper, we find no evidence of a logical negation concept not only prelinguistically, but even long after infants have begun producing the negation word “no”. Although the majority of 15-month-olds say “no”, we find that they do not comprehend the logical meaning of this word until about a year after they begin to produce it, at around 24 months. Their comprehension of logical “no” co-occurs with several related developments. At the same age, toddlers begin to produce another frequent English negation word, “not”, and to comprehend the logical meaning of “not” as well. They also consistently begin to deploy both

words to deny the truth of propositions, including both statements others have made and spontaneous, self-initiated utterances that something is not the case. The convergence of all these changes at a single age suggests a common underlying factor. That factor could be the emergence of conceptual logical negation, or an emerging ability to map a pre-existing concept to any word, either “no” or “not”. Making the mapping between the words and the logical concept depends on several aspects of language learning. Before learning that “no” and “not” negate propositions, the child must be able to represent propositions as the content of utterances, which requires knowing a sufficient number of other words as well as the syntactic and semantic rules for how these words combine. Knowing the syntactic role that a negation particle plays may also help to constrain the hypothesis space for the meanings of negation words, perhaps to the fairly small space of sentence-level semantic operators.

Given that children both fail to combine logical negation with the representation approach pre-linguistically, and fail to map it to the word “no” when they begin to produce this word, it may be tempting to conclude that they simply do not have a truth-functional negation concept. At present, however, there is a gap of a few potentially crucial months between the ages tested in these papers; we test 7- and 14-month-olds in the first paper, while the youngest group in the second paper is 20-month-olds. There is some reason to think that a compositional negation concept could emerge between these ages.

5.1.1. A suspicious coincidence

Mutual exclusivity is an inductive bias that assumes every object has just one name (Markman & Wachtel, 1988). Used as a word-learning strategy, a learner who sees some familiar referents and one novel one, and hears a novel word, would map the novel word to the novel object. This procedure might have nothing to do with negation, instead involving a domain-

specific heuristic like a preference to make novel-to-novel mappings between words and objects. But, as Halberda (2003) argues, it could also reflect a computation wherein children represent that the novel word does not map to the familiar object, and conclude that the new object is therefore more likely. Halberda (2006) presents evidence that when older children and adults are faced with the same kind of word-learning scenario, they do deploy a negation operator in just this way. Crucially, Halberda (2003) identifies 18 months as the age at which children successfully use mutual exclusivity, with the strategy emerging right between 14 and 18 months.

The last experiment in the second paper also provides suggestive evidence that a negation operator may come online around this age. In this experiment, children succeed in a search task where they are shown an empty bucket rather than being told that the ball is “not in the bucket”. In this case we find children succeeding at 20 months, significantly earlier than their success on the versions of the search task that involve “no” and “not”. There are several possible representations underlying children’s success, some of which involve negation (if children represent THE BALL IS NOT IN THE BUCKET but have not yet mapped the word “not” to the logical concept) and others that involve another domain-specific heuristic instead (like an understanding that one should avoid search in empty containers). If children start to succeed at this task between 14 and 18 months, the coincidence in having the same ages of success across a search task and a word-learning task that share nothing except a potential reliance on negation suggests a shared limiting factor.

Finally, while 14-month-olds do not appear to represent not-approach, we have not yet tested 18-month-olds. If they succeed right at the same age at which they begin to succeed on these other tasks, the coincidence would grow even greater. Although it would be possible that 18 months just happens to be the age when children develop an atomic concept avoid that is

unrelated to approach, the convergence of ages would mean that the most parsimonious explanation is one that posits a common factor developing at the same age across all three tasks – conceptual, combinatorial negation.

Given that negation can only be observed in combination with other representations, and given that there is always some positive characterization of a negated predicate, this approach is a general strategy for future research. The more different, otherwise unrelated tasks converge on the same age of success, where the only relevant feature they share is a possible reliance on conceptual negation, the more reason to think that an emerging understanding of negation underlies the common success. Of course, even if different tasks do involve a common factor, there are a variety of task-specific reasons why different tasks might not converge on the same age – limitations specific to search, or word learning, or inference about others' goals. A lack of convergence is therefore far less informative about the possible absence of negation.

5.1.2. Negation as a wedge into the conceptual system

In the introduction chapter, I argued that once an early conceptual logical negation had been identified, it could be used as a wedge into the rest of the child's conceptual system. By studying which other representations negation can combine with, we could test which of these have conceptual combinatorial properties. One issue with this strategy is that relies on the age at which negation is found being young enough. If a negation concept is late to the scene, emerging only after the child has already mastered the foundations of combinatorial language, there would already be plenty of other evidence about the combinability of different representations with each other (i.e. the child could make multi-word utterances or whole sentences). So only by acquiring evidence of a pre-linguistic or, at least an early linguistic concept of negation, would it be possible to use negation to investigate the rest of the child's conceptual repertoire. We have

not yet identified the presence of an early negation concept, but we have begun to narrow in on when and how this concept might emerge. And if the strategy of looking for converging ages of success identifies the emergence of negation at 18 months, that would be a young enough age to allow for the investigation of the conceptual status of many other representations.

For instance, children are able to represent exact quantities up to three or four, as well as larger approximate quantities at 18 months (and much earlier). However, they do not map the small numbers to number words until they are 2 to 4 years old, and the larger approximate quantities to words until 4 to 5 years old (Le Corre & Carey, 2007; see Carey, 2009 for discussion). Do these representations have semantic combinatorial properties for 18-month-olds, or, for that matter, for the 24-month-olds who the second paper has identified as having the concept of truth-functional negation? To test this, we would need to devise a mode of presenting the negation of different quantities that does not rely on number words. A don't-give-a-number task has already been used to test what 3-year-olds think number words mean before they have mapped them to the right numerical representations (e.g. "Don't give me three bananas"; Hartshorne & Barner, 2011). It may be possible to modify this task to show children the approximate quantity that they are not to give, rather than using a number word (e.g. "Don't give me this many"). If children's responses systematically exclude the negated number, we would have evidence of them combining numerical representations with logical negation. It may be possible to similarly apply this strategy in other content domains in which children can represent information at a very young age (see Carey, 2009; Spelke and Kinzler, 2007).

5.2. The development of quantification

While the first and second papers attempt to understand the development of combinatorial thought through the case study of negation, the third paper investigates

combinatorial thought in its mature state. We find that when adults process sentences that contain a scope ambiguity due to the presence of two quantifiers, the computations that assign scope to each quantifier online are distinct for the three English universals: EACH, EVERY, and ALL. We find, however, that another family of quantifiers – the numerals THREE, FOUR and FIVE – do share a scope-assigning computation, suggesting that the operation is shared by quantifiers with similar combinatorial properties but different conceptual content, but not by quantifiers with similar content but arguably different combinatorial properties.

How do children come to process these abstract relations between abstract concepts? Barner, Chow and Yang (2009) found that children begin to understand the meanings of many quantifier words between the ages of 3 and 5. One simple question that we do not yet have an answer to is what children make of scopally ambiguous sentences when they first begin to comprehend quantifier words. If children process scope ambiguities in a similar way to adults from the beginning, it would suggest that the resolution of scope ambiguity, like the phenomenon itself, is a direct consequence of the semantic relations quantifiers, not an additional compositional operation that is acquired separately. Given that some of the operations invoked to explain the resolution of scope ambiguities, and particularly the construction of inverse scope, are specialized and esoteric (e.g. quantifier-specific landing sites at LF; see Beghelli & Stowell, 1997), this is not a trivial conclusion.

Assuming children do resolve scope ambiguities in a systematic way soon after they know the meanings of the quantifier words, we could also test children's differential scope preferences to see if they match those of adults. This would provide additional evidence that 3-5 year-olds have mapped quantifier words to adult-like meanings. Priming scope resolution in children would also allow us to investigate whether the labor of conceptual combination is

divided similarly for children as for adults. The priming paradigm provided support for the distinction between conceptual content and combinatorial properties in adults. Does the same distinction exist in the conceptual repertoires of children, or do children perhaps have a single conceptual store, with the entry for each concept carrying both its meaning and instructions for how to combine it with other concepts, and with these properties differentiating at some later point in development?

5.2.1. The development of distributivity

At the end of the third paper, we suggest that the differences in scope preferences between the different universal quantifiers may be linked to a feature of quantifiers known as ‘distributivity’. Some predicates like *gather*, *be numerous*, *elect*, and *decide unanimously*, among many others, can only apply to a collection of individuals, not to any one individual (see Champollion, 2010). If a quantifier is distributive, it specifies that each individual in the quantified set has a predicate apply to it separately. Thus, a noun phrase quantified with a distributive quantifier cannot combine with a collective predicate – the predicate wants a collection, and the noun phrase provides it with individuals, resulting in a violation. Evidence that *EACH*, in particular, is a distributive quantifier comes from judgments that sentences like the following are ungrammatical: “Each boy gathered in the yard”, “Each ant in the colony was numerous”, “Each voter elected Bush” (Dowty, 1987; Champollion, 2010).

We find that the more strongly a quantifier is associated with the feature of ‘distributivity’, the stronger its preference to take wide scope. What are the developmental origins of distributivity, of its relationship to scope and to the differences between the scope-taking properties of different quantifiers? Do children understand that *EACH* is distributive and *ALL* is not from the earliest point at which they understand these quantifiers? Does the preference

for EACH to take wide scope and for ALL to take narrow scope emerge alongside children's understanding of the distributivity differences between these two quantifiers, or do children's representations of distributivity and scope preference emerge independently? Again, the developmental questions are informative above the adult representations – if the two factors emerge independently, it would suggest that the connection between distributivity and preferred scope is less principled than it seems from the results of the third paper.

5.3. Content and combination

The ability to combine concepts into propositions underlies the expressive power of both human language and thought. The papers presented in this dissertation take the representations of two kinds of logical operators, negation and quantification, as case studies of a special class of concepts – those whose content is fully determined by their combinatorial properties. These concepts serve as a direct window into the nature of combinatorial thought, both into its structure in adulthood and its developmental roots. Isolating the combinatorial properties of concepts is a useful method for studying the system of propositional thought, but a full understanding of the system will also require studying the interface between conceptual content and combination. If conceptual content and concepts' combinatorial properties are stored separately, what features of each representation enable their rapid binding in productive thought? What are the limits on which content can be bound to which combinatorial properties? Although the present work does not provide answer to these questions, it has put us in a better position to ask them.

References

- Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.
- Ameel, E., Barbara, M., & Storms, G. (2008). Object naming and later lexical development: From baby bottle to beer bottle. *Journal of Memory and Language*, 58, 262-285.
- Anderson, C. (2004). *The Structure and Real-Time Comprehension of Quantifier Scopep Ambiguity*. (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.
- Austin, K., Theakston, A., Lieven, E., & Tomasello, M. (2014). Young children's understanding of denial. *Developmental psychology*, 50(8), 2061.
- Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., Wu, D., & Bian, L. (2014). Psychological and sociomoral reasoning in infancy. In M. Mikulincer, P. R. Shaver (Eds.), E. Borgida, & J. A. Bargh (Assoc. Eds.), *APA handbook of personality and social psychology: Vol. 1. Attitudes and social cognition* (pp. 79-150). Washington, DC: American Psychological Association.
- Barker, C. (2002). Continuations and the nature of quantification. *Natural language semantics*, 10(3), 211-242.
- Barner, D., Chow, K., & Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology*, 58, 195-219.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.
- Bartoshuk, A. K. (1962). Human neonatal cardiac acceleration to sound: Habituation and dishabituation. *Perceptual and motor skills*, 15(1), 15-27.
- Bates, D., Maechler, M., Bolker, B. and Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8, <http://CRAN.R-project.org/package=lme4>.
- Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of each and every. In Szabolcsi, A. (Ed.), *Ways of scope taking* (pp. 71-107). Netherlands: Springer.
- Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental psychology*, 41(2), 328.
- Beier, J.S., & Carey, S. (2014). Contingency is not enough: Social context guides third-party attributions of intentional agency. *Developmental psychology*, 50(3), 889-902.

- Beier, J. S., & Spelke, E. S. (2012). Infants' developing understanding of social gaze. *Child development*, 83(2), 486-496.
- Biro, S., Verschoor, S., & Coenen, L. (2011). Evidence for a unitary goal concept in 12-month-old infants. *Developmental science*, 14(6), 1255-1260.
- Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.
- Bott, O., & Radó, J. (2007). Quantifying quantifier scope: a cross-methodological comparison. *Roots: Linguistics in Search of Its Evidential Base*, 96, 53-74.
- Brody, L. R., Zelazo, P. R., & Chaika, H. (1984). Habituation–dishabituation to speech in the neonate. *Developmental Psychology*, 20(1), 114.
- Bunt, H. (1985). *Mass terms and model-theoretic semantics*. Cambridge, UK: Cambridge University Press.
- Burge, T. (2010a). Steps toward origins of propositional thought. *Disputatio*, 4(29), 1-29.
- Burge, T. (2010b). *Origins of objectivity*. Oxford: Oxford University Press.
- Busby Grant, J., & Suddendorf, T. (2011). Production of temporal terms by 3-, 4-, and 5-year-old children. *Early Childhood Research Quarterly*, 26, 87–95.
- Cameron-Faulkner, T., Lieven, E., & Theakston, A. (2007). What part of no do children not understand? A usage-based account of multiword negation. *Journal of Child Language*, 34, 251–282.
- Camp, E. (2004). The generality constraint and categorial restrictions. *The philosophical quarterly*, 54(215), 209-231.
- Carey, S. (2009). *The origin of concepts*. Oxford University Press.
- Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological review*, 82(1), 45.
- Carpenter, M., Nagel, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63, 176-176.
- Catlin, J., & Micham, D. L. (1975). Semantic representations as procedures for verification. *Journal of Psycholinguistic Research*, 4, 209-225.
- Champollion, L. (2010). Parts of a whole: Distributivity as a bridge between aspect and

- measurement. (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia, PA.
- Chemla, E., & Bott, L. (2015). Using Structural Priming to Study Scopal Representations and Operations. *Linguistic Inquiry*, 46(1), 157-172.
- Choi, S. (1988). The semantic development of negation: A cross-linguistic longitudinal study. *Journal of Child Language*, 15, 517-531.
- Chomsky, N. (1976). Conditions on Rules of Grammar. *Linguistic Analysis*, 2, 303-51.
- Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive psychology*, 3(3), 472-517.
- Clark, H. H., & Chase, W. G. (1974). Perceptual coding strategies in the formation and verification of descriptions. *Memory & Cognition*, 2(1), 101-111.
- Clark, D. G., & Kar, J. (2011). Bias of quantifier scope interpretation is attenuated in normal aging and semantic dementia. *Journal of Neurolinguistics*, 24(4), 401-419
- Cooper, R. (1983). Quantification and Syntactic Theory. In D. Reidel (Ed.), *Synthese Language Library*. (Vol. 21). Dordrecht, Netherlands.
- Cushman, F., Murray, D., Gordon-McKeon, S., Wharton, S., & Greene, J. D. (2011). Judgment before principle: engagement of the frontoparietal control network in condemning harms of omission. *Social cognitive and affective neuroscience*, 7(8), 888-895.
- Cushman, F. A. & Young, L. (2011). Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, 35: 1052-1075.
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28(1), 125-127.
- Dehaene, S. (1997). *The number sense*. New York: Oxford University Press.
- de Villiers, J. G., & Flusberg, H. B. T. (1975). Some facts one simply cannot deny. *Journal of Child Language*, 2(02), 279-286.
- Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and knowledge* (pp. 67-110). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, 74(1), 24-40.

- Donaldson, M. (1970). Developmental aspects of performance with negatives. In G.B. Flores D'Arcais & W.J.M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam & London: North-Holland.
- Dowty, D. (1987). A note on collective predicates, distributive predicates, and 'all'. In F. Marshall (Ed.), *Proceedings of the Third Eastern States Conference on Linguistics (ESCOL 86)*. The Ohio State University, Columbus, OH. (pp. 97-115).
- Egyed, K., Király, I., & Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychological Science*, 24(7), 1348-1353.
- Evans, G. (1982). *The varieties of reference*. J. McDowell (Ed.). Oxford: Clarendon Press.
- Feiman, R., Carey, S., & Cushman, F. (2015). Infants' representations of others' goals: Representing approach over avoidance. *Cognition*, 136, 204-214.
- Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, 21(01), 95-116.
- Filik, R., Paterson, K. B., & Liversedge, S. P. (2004). Processing doubly quantified sentences: Evidence from eye movements. *Psychonomic Bulletin & Review*, 11(5), 953-959.
- Fodor, J. A. (1975). *The language of thought*. Harvard University Press.
- Fodor, J. D. (1982). The mental representation of quantifiers. In S. Peters & E. Saarinen (Eds.), *Processes, beliefs, and questions: Essays on formal semantics of natural language and natural language processing*. 129-164. Dordrecht: Reidel.
- Fodor, J. D., & Sag, I. (1982). Referential and quantificational indefinites. *Linguistics & Philosophy*, 5, 355-398.
- Gallistel, C.R., & King, A. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience*. New York: Blackwell/Wiley.
- Geach, P. T. (1972). History of a Fallacy. *Logic matters*, 1-13.
- Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naive theory of rational action. *Trends in Cognitive Science*, 7, 287-292.
- Gillen, K. (1991). *The comprehension of doubly quantified sentences*. (Unpublished doctoral dissertation), University of Durham, UK.

- Goldin-Meadow, S., Seligman, M., & Gelman, R. (1976). Language in the two-year old. *Cognition*, 4, 189-202.
- Gredebäck, G., Stasiewicz, D., Falck-Ytter, T., Rosander, K., & von Hofsten, C. (2009). Action type and goal type modulate goal-directed gaze shifts in 14-month-old infants. *Developmental psychology*, 45(4), 1190.
- Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.
- Guidetti, M. (2005). Yes or no? How do young children combine gestures and words to agree and refuse. *Journal of Child Language*, 32, 911–924.
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), 23-34.
- Hamlin, J.K., & Wynn, K. (2012). Who knows what's good to eat? Infants fail to match the food preferences of antisocial others. *Cognitive Development*, 27, 227-239.
- Hamlin, J. K., Wynn, K., Bloom, P. (2010). 3-month-olds show a negativity bias in their social evaluations. *Developmental Science*, 13(6), 923-929.
- Harris, M., Yeeles, C., Chasin, J. & Oakley, Y. (1995). Symmetries and asymmetries in early lexical comprehension and production. *Journal of Child Language*, 22, 1-18.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar* (Vol. 13). Oxford: Blackwell.
- Hendriks, H. (1988). Type Change in Semantics: the Scope of Quantification and Coordination. In E. Klein and J. van Benthem, (Eds.), *Categories, Polymorphism and Unification*, ITLI, Amsterdam, (pp. 96–119).
- Hernik, M., & Southgate, V. (2012). Nine-months-old infants do not need to know what the agent prefers in order to reason about its goals: on the role of preference and persistence in infants' goal-attribution. *Developmental science*, 15(5), 714-722.
- Hornik, R., Risenhoover, N., & Gunnar, M. (1987). The effects of maternal positive, neutral, and negative affective communications on infant responses to new toys. *Child Development*, 937-944.
- Hornstein, N. (1984). *Logic as grammar*. MIT Press.
- Hummer, P., Wimmer, H., & Antes, G. (1993). On the origins of denial negation. *Journal of Child Language*, 20, 607–618.
- Ioup, G. (1975). Some universals for quantifier scope. In J. Kimball (Ed.), *Syntax and Semantics*, 4, 37-58. New York: Academic Press.
- Jackendoff, R. (1981). On Katz's autonomous semantics. *Language*, 57, 425-435.

- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434-446.
- Johnson-Laird, P. N. (1969). On understanding logically complex sentences. *Quarterly Journal of Experimental Psychology*, 21, 1-13.
- Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1), 67-81.
- Katz, J. J. (1980). *Propositional structure and illocutionary force: A study of the contribution of sentence meaning to speech acts* (Vol. 6). Harvard University Press.
- Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38(7), 1033-1050.
- Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2006). Experiential simulations of negated text information. *The Quarterly Journal of Experimental Psychology*, 60, 976-990.
- Keller, W.R. (1988). Nested Cooper storage: The proper treatment of quantification in ordinary noun phrases. In U. Reyle & C. Rohrer (Eds.), *Natural language parsing and linguistic theories*, 432-447. Dordrecht: Reidel.
- Kim, K. J. (1985). Development of the concept of truth-functional negation. *Developmental Psychology*, 21, 462-472.
- Kosslyn, S. M. (1980). *Image and mind*. Harvard University Press.
- Kowalski & Zimiles (2006). The relation between children's conceptual functioning with color and color term acquisition. *Journal of Experimental Child Psychology*, 94, 301-321.
- Kroch, A. (1974). *The semantics of scope in English*. (Unpublished doctoral dissertation), MIT, Cambridge, MA.
- Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, 48, 243-279.
- Lakoff, G. (1971). On generative semantics. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*, (pp. 232-296). Cambridge: Cambridge University Press.
- Lakusta, L., Wagner, L., O'Hearn, K., & Landau, B. (2007). Conceptual foundations of spatial language: Evidence for a goal bias in infants. *Language learning and development*, 3(3), 179-197.

- Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105(2), 395-438.
- Lidz, J., & Musolino, J. (2002). Children's command of quantification. *Cognition*, 84(2), 113-154.
- Lloyd, P., & Donaldson, M. (1976). On a method of eliciting true/false judgments from young children. *Journal of Child Language*, 3(03), 411-416.
- Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? *Psychological Science*, 16(8), 601-608.
- Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled object. *Cognitive Psychology*, 58(4), 441-486.
- Luo, Y., & Johnson, S.C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, 12(1), 142-149.
- May, R. (1977). *The Grammar of Quantification*. (Doctoral dissertation), MIT, Cambridge, MA.
- May, R. (1985). *Logical form: It's structure and derivation*. Cambridge, MA: MIT Press.
- Micham, D. L., Catlin, J., VanDerveer, N. J., & Loveland, K. A. (1980). Lexical and structural cues to quantifier scope relations. *Journal of Psycholinguistic Research*, 9, 367-377.
- Moll, H. (2013). Ontogenetic precursors of assertion and denial. In S. Roedl, & H. Tegtmeyer (Eds.), *Essays in honor of Pirmin Stekeler-Weithofer* (pp. 337-346). Boston, MA: DeGruyter.
- Montague, R. (1970). English as Formal Language. in B. Visentini, et al. (Eds), *Linguaggi nella Società e nella Tecnica* (Edizioni di Comunità), (pp. 189-224).
- Montague, R. (1972). The Proper Treatment of Quantification in Ordinary English. in J. Hinrikka, J. Moravcsik, & P. Suppes (Eds.), *Approaches to Natural Language*, (pp. 221-242).
- Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, 61(2), 63-86.
- Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle an event-related potential study on the pragmatics of negation. *Psychological Science*, 19(12), 1213-1218.

- Nordmeyer, A. E., & Frank, M. C. (2014). The role of context in young children's comprehension of negation. *Journal of Memory and Language*, 77, 25-39.
- Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science*, 17(9), 781-787.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255-258.
- Pea, R. D. (1980a). The development of negation in early child language. In D. R. Olson (Ed.), *The social foundations of language and thought: Essays in honor of Jerome S. Bruner* (pp. 156–186). New York, NY: W. W. Norton.
- Pea, R. D. (1980b). Logic in early child language. *Annals of the New York Academy of Sciences*, 345(1), 27-43.
- Pea, R. D. (1982). Origins of verbal logic: Spontaneous denials by two-and three-year-olds. *Journal of Child Language*, 9(3), 597-626.
- Peacocke, C. (1992). *A Study of Concepts*, Cambridge MA: MIT Press.
- Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, 85(1), 53-78.
- Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4), 633-651.
- Pinsker, H.M., Kupferman, I., Castellucci, V.F., & Kandel, E.R. (1970). Habituation and dishabituation of the gill-withdrawal reflex in *Aplysia*. *Science*, 167, 1740-1742.
- Pitchford, N., & Mullen, K. (2003). The development of conceptual colour categories in pre-school children: Influence of perception on categorization. *Visual Cognition*, 10, 51–77.
- Quine, W.V.O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Raffray, C. N., & Pickering, M. J. (2010). How do people construct logical form during language comprehension?. *Psychological science*, 21(8), 1090-1097.
- Reinhart, T. (1997). Quantifier scope: how labor is divided between QR and choice functions. *Linguistics and Philosophy*, 20, 335–397.
- Repacholi, B.M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds.

- Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes*, 79(2), 79-94.
- Roberts, C. (1987). *Modal Subordination, Anaphora, and Distributivity*. Doctoral dissertation. University of Massachusetts at Amherst, MA.
- Robson, S. J., & Kuhlmeier, V. A. (2013). Selectivity promotes 9-month-old infants to encode the goals of others. Poster presented at the meeting of the Society for Research in Child Development. Seattle, WA.
- Rodman, R. (1976). Scope Phenomena, 'Movement Transformations', and Montague Grammar. In B. H. Partee (Ed.), *Montague Grammar*. New York: Academic Press. (pp. 165-176).
- Ruys, E., & Winter, Y. (2011). Scope ambiguities in formal syntax and semantics. In D. Gabbay & F. Gunthner (Eds.), *Handbook of Philosophical Logic* (2nd Ed). Amsterdam: John Benjamins. (pp. 159-225).
- Sandhofer, C., & Smith, L. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology*, 35, 668–679.
- Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret Agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science*, 16(12), 995-1001.
- Saxe, R., Tzelnic, T., & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology*, 43(1), 149-158.
- Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, 18(9), 668-671.
- Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological science*, 22(7), 878-880.
- Shatz, M., Tare, M., Nguyen, S. P., & Young, T. (2010). Acquiring non-object terms: The case for time words. *Journal of Cognition and Development*, 11, 16–36.
- Snedeker, J., Geren, J., & Shafto, C. (2012). Disentangling the effects of cognitive development and linguistic expertise: A longitudinal study of the acquisition of English in internationally-adopted children. *Cognitive Psychology*, 65, 39-76
- Snedeker, J., Geren, J., & Shafto, C. (2007). Starting over: International adoption as a natural experiment in language development. *Psychological Science*, 18(1), 79-87.
- Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon* (pp. 257-294). Cambridge, MA: MIT Press.

- Sodian, B., & Thoermer, C. (2004). Infants' understanding of looking, pointing, and reaching as cues to goal-directed action. *Journal of Cognition and Development*, 5(3), 289-316.
- Soja, N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development*, 65, 918-937.
- Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, 96(1), B1-B11.
- Spaepen, E., & Spelke, E. (2007). Will any doll do? 12-month-olds' reasoning about goal objects. *Cognitive Psychology*, 54, 133-154.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, 10(1), 89-96.
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27(1), 76-105.
- Steedman, M. (2012). *Taking scope: The natural semantics of quantifiers*. Cambridge, MA: MIT Press.
- Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology*, 63(12), 2305-2312.
- Tillman, K., & Barner, D. (2015). Learning the language of time: Children's acquisition of duration words. *Cognitive Psychology*, 78, **57-77**.
- Tunstall, S. L. (1998). *The interpretation of quantifiers: Semantics and processing*. (Unpublished doctoral dissertation), University of Massachusetts, Amherst, MA.
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological bulletin*, 134(3), 383.
- Van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of memory and language*, 41(2), 147-182.
- VanLehn, K. (1978). *Determining the scope of English quantifiers*. (Unpublished Master's Thesis), MIT, Cambridge, MA.
- Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press.
- von Stechow, A. (1991). Syntax and Semantics. in A. von Stechow & D. Wunderlich

- (Eds.), *Semantik/Semantics. An International Handbook of Contemporary Research*. Berlin and New York: de Gruyter. (pp. 90-148).
- Vul, E., & Kanwisher, N. (2010). Begging the question: The non-independence error in fMRI data analysis. *Foundational issues for human brain mapping*, 71-91.
- Wagner, K., Dobkins, K & Barner, D. (2013). Slow mapping: Color word learning as a gradual inductive process. *Cognition*, 127, 307-317.
- Wason, P. C. (1965). The contexts of plausible denial. *Journal of verbal learning and verbal behavior*, 4(1), 7-11.
- Widen, S. & Russell, J. (2003). A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, 39, 114-128.
- Winter, Y. (2001). *Flexibility principles in Boolean semantics: The interpretation of coordination, plurality, and scope in natural language*. Boston MA: MIT press.
- Woodward, A.L., & Guajardo, J.J. (2002). Infants' understanding of the point gesture as an object-directed action. *Cognitive Development*, 17, 1061–1084.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1-34.
- Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant behavior and development*, 22(2), 145-160.
- Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, 6(3), 297-311.
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, 24, 220–251.